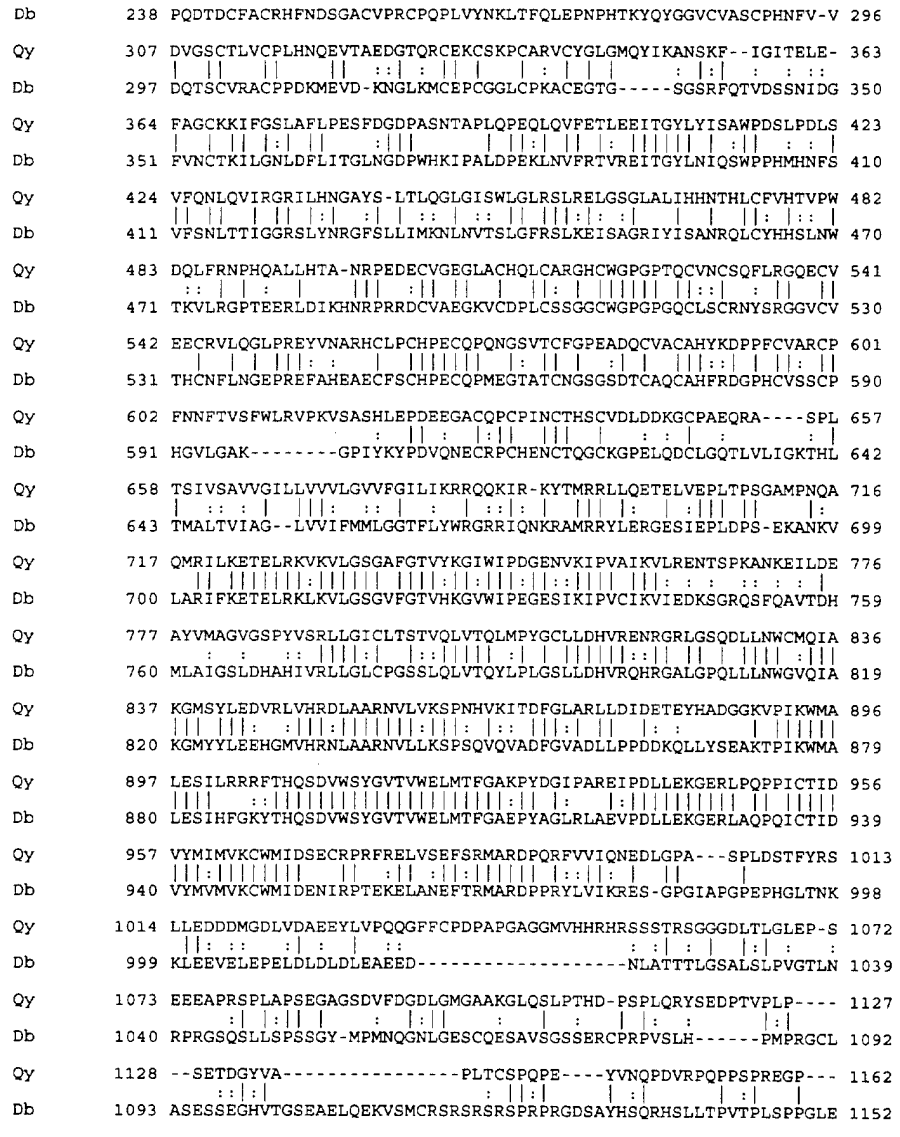
```
RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Gallus gallus (chicken)
C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C;Accession: A27720; A00643
R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A;Reference number: A27720; MUID:88261272; PMID:3260329
A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and prc
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
pecific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TMM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

  Query Match             44.6%;  Score 3053.5;  DB 1;  Length 1223;
  Best Local Similarity   48.2%;  Pred. No. 1.7e-119;
  Matches  631;  Conservative  169;  Mismatches  347;  Indels  161;  Gaps   29;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |   :| ||  | |        :|| ||: ||      | |    |: :|  |:||  |||
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  |:||: |||| |||||| | |  :||: |:|:||  |:::::||||| |
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
               :| |          ||||| :: |:||| ||| |  ||:||  ||:|| ||    : :  |
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              |::|  :   :|  | | |     |||   ::||:||: :||  |: ||:| :|:|||| |
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: ||||||  |      |:  || || ||  |::   ||||:|:|||:||  |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||   |      ||  |:| ::|:||   |::|| |:|:  :|     |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  | || | :: || :| ||  : | || |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
              ||  |:||::|||   :| ||| : | |  ||||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: |  ||     |       || :::|  :   |  ||:   ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
              |||::| :||| |||:     ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAE 651
              ||  ||      | |: |                |    || |  ||| |     :|||
Db        601 CVKACPAGVLGENDTL-VWKYA----------DANAVCQLCHPNCTRGCKGPGLEGCP-- 647

Qy        652 QRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSG 710
              |   || : || |:| :||:|:  |: ::||   :|| |:|||||| ||||||||||
Db        648 -NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSG 705

Qy        711 AMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKAN 770
                |||| :||||||| :|||||||||||||||||:|||:|| |||||||| ||| ||||||
Db        706 EAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKAN 765

Qy        771 KEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLN 830
              |||||||||| | :|:| |||||||||||||:|||||||||||::||::  :||| |||
Db        766 KEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLN 825

Qy        831 WCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKV 890
              ||:|||||:|||: |||||||||||||:| |||||||||:||  || ||||:||||
Db        826 WCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKV 885

Qy        891 PIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQP 950
              |||||||||| | :|||||||||||||||||||:|||||||| ||  :|||||||||
Db        886 PIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQP 945

Qy        951 PICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDST 1009
              |||||||||||||||:: ||:||||::|||:|||| |::||| :| :   || ||
Db        946 PICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSK 1005

Qy       1010 FYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGL 1069
              |||:|:|::|| |:|||:|||| ||||                  | ||
Db       1006 FYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST---------- 1038

Qy       1070 EPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV 1124
                 |:||     | |  : ::  |       :  | |  : | :|||| |||
Db       1039 -------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTG 1083

Qy       1125 PLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPG 1182
                |:  ||::      | ||||||  : |: ||                     :
Db       1084 NFLEESIDDGFL------PAPEYVNQ--LMPKKPS--------------------TAMV 1114

Qy       1183 KNGVVKDVF--------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLY 1227
              :| :  ::                :   ||:|||||          :  |     |::
Db       1115 QNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFESSP 1166

Qy       1228 YWDQ--------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1260
              || |        | |:          |:     | ||||||  :  |
Db       1167 YWIQSGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 1214


RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide
```

```
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

  Query Match           43.1%;  Score 2945.5;  DB 2;  Length 1308;
  Best Local Similarity  44.9%;  Pred. No. 5.3e-115;
  Matches  609;  Conservative  181;  Mismatches  378;  Indels  189;  Gaps   30;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  | ||||: ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::|||| ||| : :|  ||| ||| |:||          |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||    ||  ||:   |   :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||    ||  |:| : |: |:  | : | |:|   : :          :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||    ||| :   : ||:| || |: ||||:| | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||  | ||:|::|    | |:  ||: |  :||:|: :|    |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||   :| ::   ||  : |  ||: |: ||:   |||||| ||::| :| ||: |:|
Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCP- 601
              | :  |  ||: |    |: | |:|:  ::| :|| ||  | |  |:|:|| | || :||
Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        602 ----FNNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGC--------- 648
                  |:|  :          :|| |  | ||  |||  |          |
Db        596 GLQGANSFIFKY----------ADPDRE--CHPCHPNCTQGCNGPTSHDCIYYPWTGHS 642

Qy        649 --PAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEP 705
                |  | :||  | : |: |: ::|::|: | : ::|:  | :|  :|| | |||||||
Db        643 TLPQHAR-TPL--IAAGVIGGLFILVIVGLTFAVYVRRKSIK-KKRALRRFL-ETELVEP 697

Qy        706 LTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENT 765
              |||||  |||||:||||||::|||||||||||||||:|:|| |||||||:| | |
Db        698 LTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETT 757

Qy        766 SPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGS 825
              ||||  | :||| :||  :  |:: ||||:||: |:|||||||:||||::|  |::  :||
Db        758 GPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGS 817

Qy        826 QDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHA 885
              | |||||:|||||  |||: |||||||||||||||||||||||||||||||||||: || ||:|
Db        818 QLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNA 877

Qy        886 DGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGE 945
              ||||:|||||||| |  |:|||||||||||:|||||| |||||| ||||||||||||||
Db        878 DGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGE 937

Qy        946 RLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPAS 1004
              |||||||||||||:|||||||:: ||:|:|| :||||||||||::||| :| :   |
Db        938 RLPQPPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPS 997

Qy       1005 PLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGD 1064
              | || |:::||:::|: |::|||||||| | |  | |       :   | |  | ||
Db        998 PNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS---- 1046

Qy       1065 LTLGLEPSEEEAPRS---------------------------PLAP-SEGAGSDVFDGD 1095
                :|  |     | |                          | || ::|| :::||
Db       1047 -EIGHSPPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDS 1105

Qy       1096 LGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQ 1148
                |  :   :   : |  |||| |||| |         : :||: |:   |: ||:|
Db       1106 CCNGTLRKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNP 1165

Qy       1149 PDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGG 1208
                :          | |  : |                ||| ::       |::|||     |
Db       1166 VE---------ENPFVSRR--------------KNGDLQ-------ALDNPEYHNASNG 1194

Qy       1209 AAPQPHPPPA---------------------------------FSPAFDNLYYWDQDPP 1234
                  || |                                    ||||  ||:   |
Db       1195 ------PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLP 1248

Qy       1235 ERGA--PPSTFKGTPT-------------AENPEYL 1255
                |     |   :  |             |||||||
Db       1249 PRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transfo
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueler, W.; Raulf, F.; Telling, A.; Ro
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing 1
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueler, W.; Schartl, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025,'N',1027-1098,'A',1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

  Query Match           38.4%;  Score 2627;  DB 1;  Length 1166;
  Best Local Similarity  44.5%;  Pred. No. 7.5e-102;
```

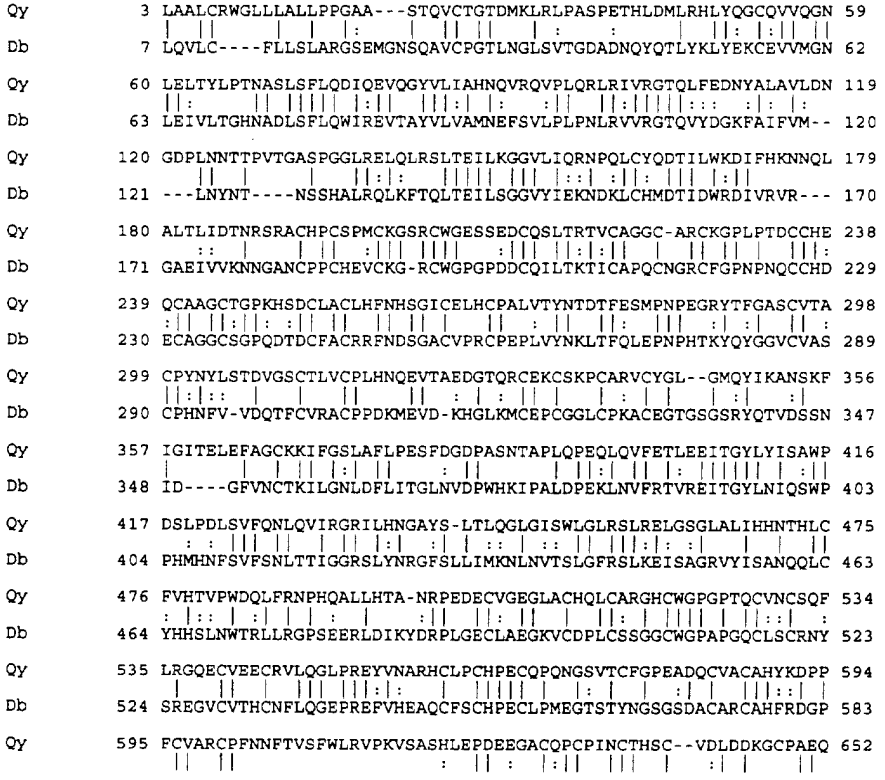
```
  Matches  568;  Conservative  165;  Mismatches  392;  Indels  152;  Gaps   31;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              ||| :  ||| |       ||    :|| ||  :: :    :  |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||   |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: | |: |
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                | ::|      ||::||| :||||| ||| :  || ||  :|| | ||  | :
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
              : ||      | |  |    |    ||       ||  |: :||  |  ||:|| | |||:|
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              || |||||: :|||||  ||  | |:  ||     |:  : : :  ||  :|||||:||
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
              || ||: |: |:|    |      ||  |:| : |:|   | :|| |:|:  : :|:  :
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
              | :  |: | || | :      ||:|||      : || |    |::|||||  |  ||:
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
              ::  |||||||::||||   :  :| : :|   : |||||||:|: :|  :: :   | :
Db        417 NMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRY 476

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
              :|: | :|||:  |:: : |          |   |:  || |||| ||:|    |
Db        477 ANTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCW-PGPTMCVSCLHVDR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFC 596
              |  ||  | :||| |||       |: || ||  |  |:||:||    |   ||::| | |
Db        529 GGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKSAHFQDGPQC 588

Qy        597 VARCP---FNNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAEQR 653
              : |||     :    |                |: | ||||  |||  |      ||  :
Db        589 IPRCPHGILGDGDTLIWKYA----------DKMGQCQPCHQNCTQGCSGPGLSGCRGD-I 637

Qy        654 ASPLTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMP 713
              |   :  |  | |:|: |:: ::  :|::||: | || |:| |||  ||||||||| |
Db        638 VSHSSLAVGLVSGLLITVIVALLIVVLLRRRRIK-RKRTIRCLLQEKELVEPLTPSGQAP 696

Qy        714 NQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEI 773
              ||| :|||||||  :| :|||||||||||||:| |||||::|||||||||| |||| |:|:
Db        697 NQAFLRILKETEFKKDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQEV 756

Qy        774 LDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCM 833
              |||||||  |  |:| |||||||| ||||||||||||||||:||::: |:  | |||||:
Db        757 LDEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCV 816

Qy        834 QIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIK 893
              ||||||:|||:  |||||||||||:|:|||||||||::||  || || |||||||||
Db        817 QIAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPIK 876

Qy        894 WMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPIC 953
              ||||||||:  :|||||||||||||||||||:|||||||:||  :|| |||||||||
Db        877 WMALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPIC 936

Qy        954 TIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRS 1013
              ||:||||::||||||   |||||||| |||:|||| |::|||       |  |   :
Db        937 TIEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSLSDRRLFSR 993

Qy       1014 LLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSE 1073
              ||  ||  |:|||:|||:| :                    | |
Db        994 LLSSDD--DVVDADEYLLPYKRI--------------------NRQGS---------- 1019

Qy       1074 EEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDG 1132
                    |  |  |                        |  : |   |   |||   |  : ||
Db       1020 -----EPCIPPTGH--------------------PVRENSITLRNISDPTQNALEKDLDG 1054

Qy       1133 YVAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKTL 1179
              :          ||||||                |:  :     ||:  ::: |      | : |
Db       1055 H----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYL 1104

Qy       1180 SPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAP 1239
              :  :| :          |::::|:|   | |           :  ||          |: ||
Db       1105 NTNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGAL 1139

Qy       1240 PSTFKGTPTAENPEYLG 1256
                      | ||| ||||
Db       1140 TGNGMFLPAAENLEYLG 1156


RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C;Accession: A36223; I59164
R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epider
A;Reference number: A36223; MUID:90083234; PMID:2687875
A;Accession: A36223
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1342 <KRA>
A;Cross-references: GB:M29366
R;Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todarc
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A;Title: Molecular cloning and expression of another epidermal growth factor recept
A;Reference number: I59164; MUID:90311312; PMID:2164210
A;Accession: I59164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-559,'G',561-957,'F',959-1063,'G',1065-1342 <RES>
A;Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C;Genetics:
A;Gene: GDB:ERBB3; HER3
A;Cross-references: GDB:119880; OMIM:190151
A;Map position: 12q13-12q13
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; phosphotransferase
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

  Query Match             34.8%;  Score 2378.5;  DB 2;  Length 1342;
  Best Local Similarity   39.8%;  Pred. No. 1.7e-91;
  Matches  526;  Conservative  193;  Mismatches  460;  Indels  141;  Gaps   34;

Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
              ||| :|         |  || ||   | :    |      |  ||: |:|| ||||:
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
              || ||||| |:|| ||||:| |:   :||  ||:|||||:::  :|: |:     ||  |
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
                |   ||:|:|  ||||| ||| |::| :||: ||| |:||   :   :: :
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
              |:| ||  :||| |||  |||||:|:|:||  |   | || |  |||::|| ||:|
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
              |: :|| || ||| || |   ||  : ||  ||:  |||  :| :|  || :||:|::
```

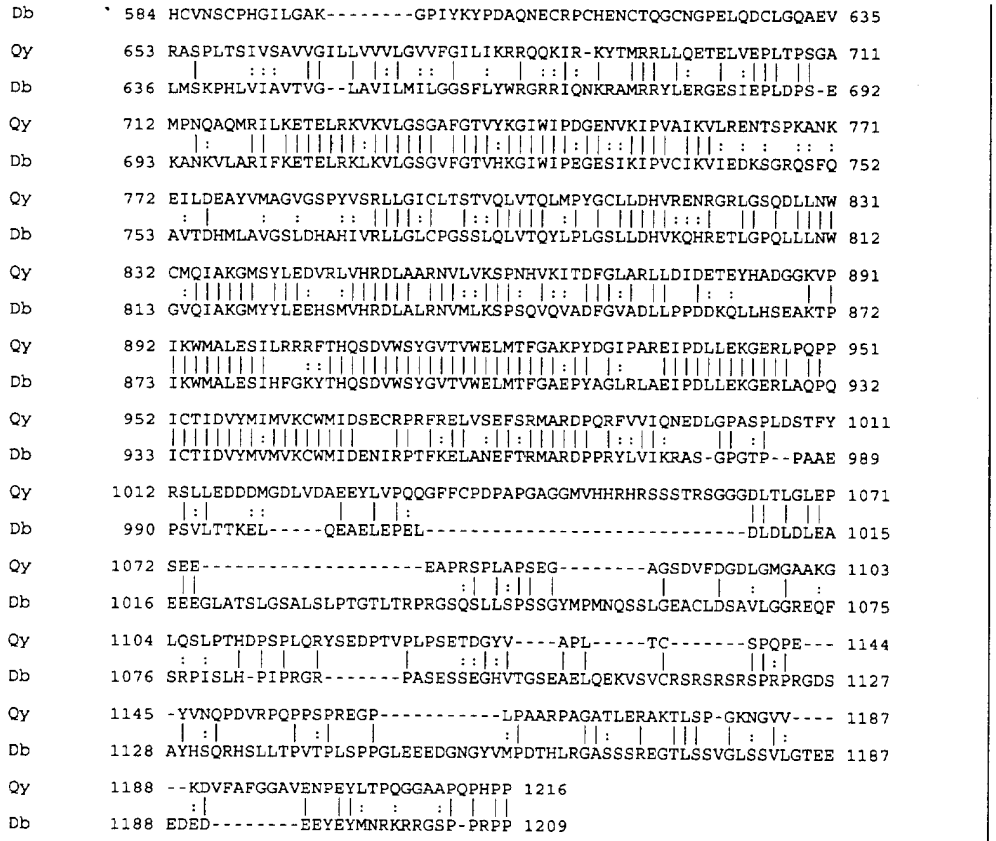


RESULT 10
JC4387

epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant pro
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for re
C;Comment: This protein is a functional heregulin receptor that transduces signals
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prot
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TMM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (T

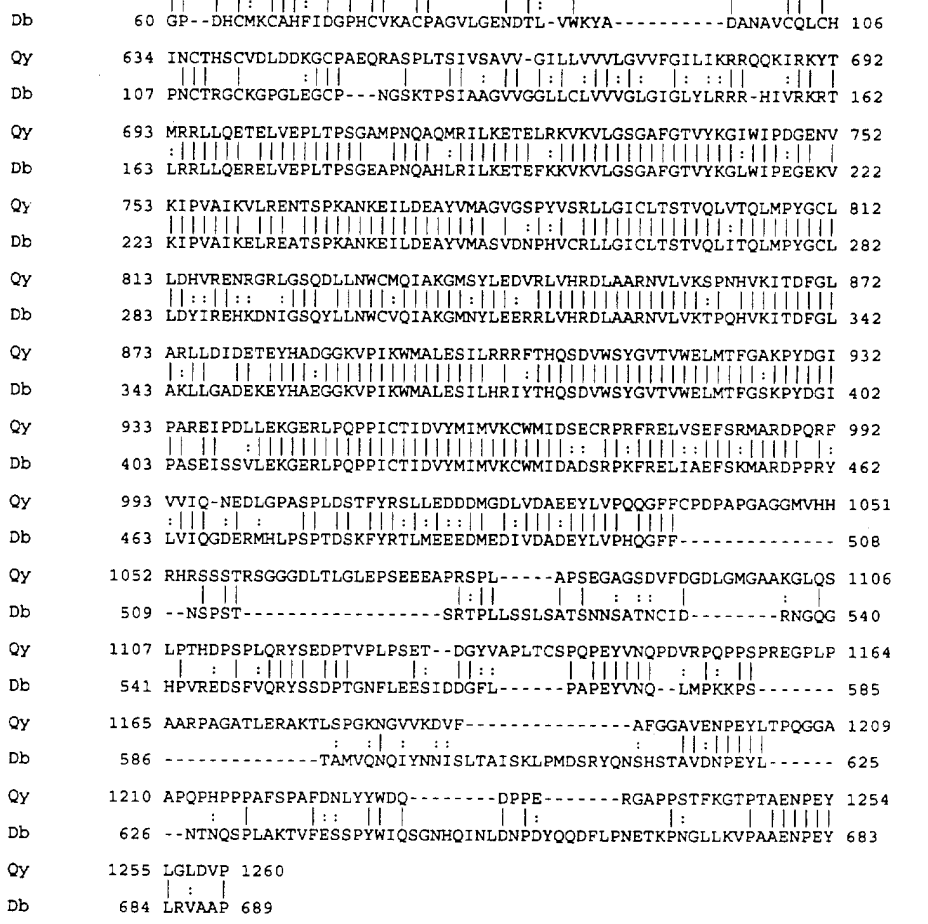
Query Match 33.5%; Score 2293.5; DB 2; Length 1339;
Best Local Similarity 40.1%; Pred. No. 5.7e-88;
Matches 518; Conservative 168; Mismatches 440; Indels 165; Gaps 35;





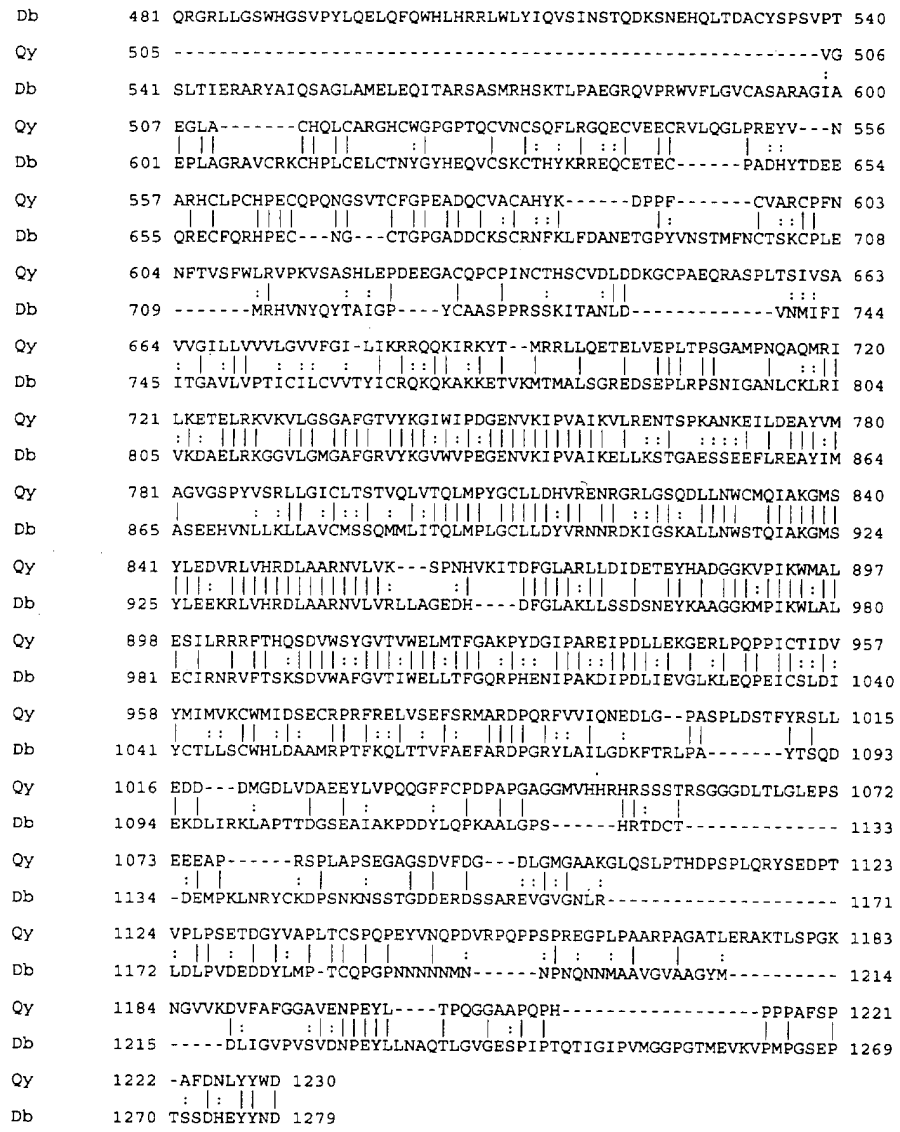
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and prc
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.2%; Score 1725.5; DB 1; Length 698;
Best Local Similarity 51.4%; Pred. No. 1e-64;
Matches 373; Conservative 76; Mismatches 138; Indels 139; Gaps 20;



RESULT 12
TVYUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678

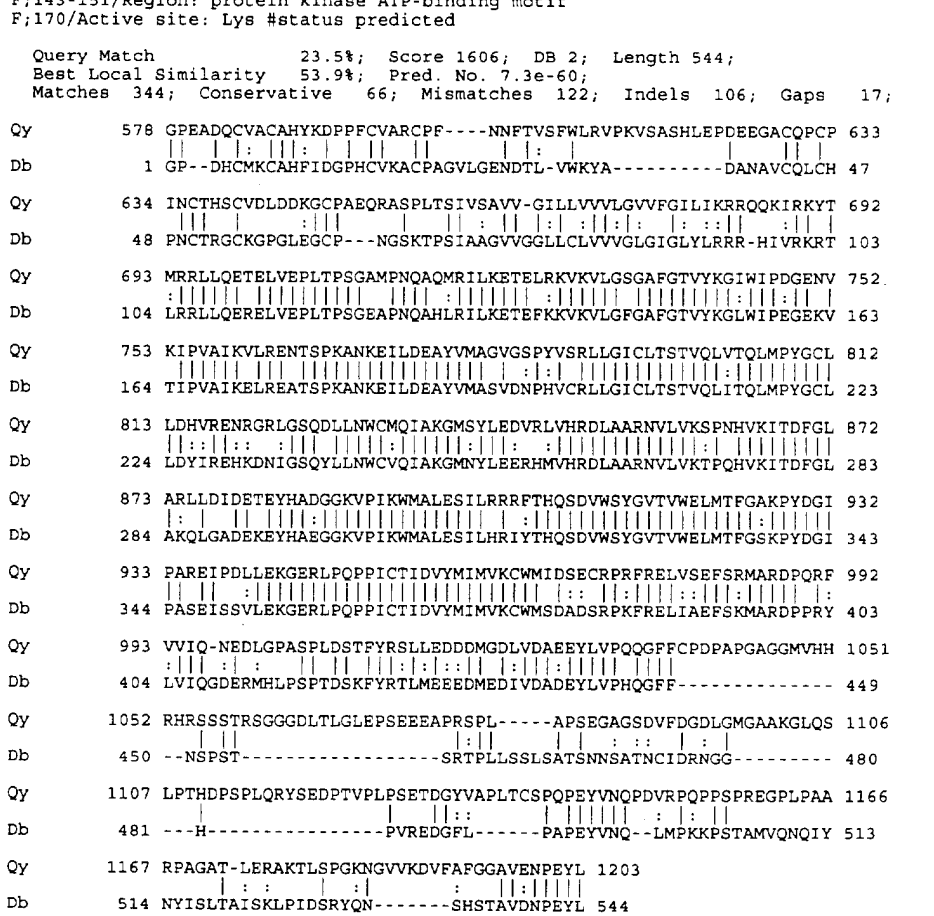
```
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

  Query Match             24.5%;  Score 1677;  DB 1;  Length 604;
  Best Local Similarity   51.6%;  Pred. No. 9.3e-63;
  Matches  361;  Conservative   73;  Mismatches  126;  Indels  140;  Gaps   18;

Qy        587 CAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVD 642
              |||: | | ||  ||      | |:     | | :      |    || |  |||  |
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-----VRKYA------DANAVCQLCHPNCTRGCKG 51

Qy        643 LDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETE 701
                :|||      |   || : || |:| :||:|:  |: ::||   :|| |:|||||| |
Db         52 PGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERE 107

Qy        702 LVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVL 761
              |||||||||  |||| :|||||| :|||||||||||:|||:|||:|| |||||||| |
Db        108 LVEPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKEL 167

Qy        762 RENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRG 821
              || |||||||||||||||| | :|:| ||||||||||||:|||||||||||::||::
Db        168 REATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKD 227

Qy        822 RLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDET 881
               :||| |||||:||||||:|||: ||||||||||||||:| ||||||||||:||   ||
Db        228 NIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEK 287

Qy        882 EYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLL 941
              ||||:|||||||||||||| | :|||||||||||||||||||:|||||||| ||   :|
Db        288 EYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVL 347

Qy        942 EKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDL 1000
              |||||||||||||||||||||||||:: ||:||||::|||:|||||  |::||| :| :
Db        348 EKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERM 407

Qy       1001 GPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRS 1060
                || || |||:|:|::|| |:|||:|||| ||||                 | ||
Db        408 HLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-- 449

Qy       1061 GGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPL 1115
                                |:||     |.|  : ::  |         :  |  |  : | :
Db        450 ----------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFV 485

Qy       1116 QRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATL 1173
              |||| |||     |:  ||::      | ||||||  : |: ||
Db        486 QRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPSTAM------------ 524

Qy       1174 ERAKTLSPGKNGVVKDVFAF-------------------GGAVENPEYLTPQGGAAPQPH 1214
                           |   :: |                    ||:|||||          :
Db        525 -----------VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTN 564

Qy       1215 PPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254
                |      |::  || |    :             :||:|
Db        565 QSPLAKTVFESSPYWIQSGNHQ----------INLDNPDY 594


RESULT 13
GQFFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone bir
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <LIV>
A;Cross-references: EMBL:K03054
R;Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth fact
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021
A;Molecule type: DNA
A;Residues: 'A',832-866,'V',868-943,'QTPSLVK' <WAD>
A;Cross-references: EMBL:X02293; NID:g7922; PIDN:CAA26157.1; PID:g929565
C;Comment: This sequence is tentative because the introns have not been identified
C;Genetics:
A;Gene: FlyBase:Egfr
A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2 57F
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TMM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

  Query Match             23.8%;  Score 1625.5;  DB 1;  Length 1330;
  Best Local Similarity   29.6%;  Pred. No. 2.7e-60;
  Matches  412;  Conservative  177;  Mismatches  414;  Indels  387;  Gaps   40;

Qy         80 VQGYVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASP 134
              :  |::|  : :        ||:|:||  ||      |: ||| |             :
Db         38 ITNYIVIGLDLIPCTLSYRLQIIRGRTLFSLSVEEEKYALFV----------------TY 81

Qy        135 GGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHP 194
                :  |::  | ::| | |    |  ||:  || | :|                 | | |
Db         82 SKMYTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPK 141

Qy        195 CSPMCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDC 252
              |   |    ||||  ::||  ::  |:  ||  || || | :|||  || |||||   ||
Db        142 CHESCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDC 200

Qy        253 LACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCT 312
              :|| :|    : :  || :  ||  |:    ||||:| :||:||  ||  :|  | |:|
Db        201 IACKNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACV 259

Qy        313 LVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAG 366
                ||    :   |      |  |: || : |                |:| |       |
Db        260 RSCPQDKMDKGGE-----CVPCNGPCPKTC-------------PGVTVLHAGNIDSFRN 300

Qy        367 CKKIFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDS 418
              |  | |::  | ::| |  |  :|        || ||: :|| |::|||||| |
Db        301 CTVIDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQ 360

Qy        419 LPDLSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFV 477
              :|| |:||: | || |    : :| :    :  | :|:|::: ||  :| ||  ||:|
Db        361 FRNLSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYV 420

Qy        478 HTVPWDQLFRNPHQALLHTANRPEDEC---------------------------------- 504
              : |  : : | |  :     |   | |
Db        421 SNIRWPAIQKEPEQKVWVNENLRADLCGKFLTILISVQHNIIMHIFAICREKWNHLLGSV 480

Qy        505 ---------------------------------------------------------- 504
```



RESULT 14
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
R;Vennstroem, B.
submitted to the EMBL Data Library, March 1993
A;Reference number: S35743

A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X12707
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-spec
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 23.5%; Score 1606; DB 2; Length 544;
Best Local Similarity 53.9%; Pred. No. 7.3e-60;
Matches 344; Conservative 66; Mismatches 122; Indels 106; Gaps 17;



RESULT 15
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus n
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA

```
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

  Query Match             23.4%;  Score 1599;  DB 2;  Length 545;
  Best Local Similarity   53.9%;  Pred. No. 1.4e-59;
  Matches  344;  Conservative   65;  Mismatches  123;  Indels  106;  Gaps   17;

Qy        578 GPEADQCVACAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCP 633
              ||  | |: |||: | | ||  ||      | |:  |              |    || |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYA----------DANAVCQLCH 47

Qy        634 INCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYT 692
               |||  |     :|||     |   || : || |:| :||:|:  |: ::||   :|| |
Db         48 PNCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRT 103

Qy        693 MRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENV 752
              :|||||| |||||||||| |||| :||||||| :|||||| |||||||||:|||:|| |
Db        104 LRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKV 163

Qy        753 KIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCL 812
              |||||| ||| |||||||||||||||| | :|:| ||||||||||||:|||||||||
Db        164 TIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCL 223

Qy        813 LDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGL 872
              ||::||::  :||| |||||:|||||:|||:  |||||||||||||:|  ||||||||
Db        224 LDYIREHKDNIGSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGL 283

Qy        873 ARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 932
              |: |  || |||:|||||||||||||| | :|||||||||||||||||||:||||||
Db        284 AKQLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGI 343

Qy        933 PAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRF 992
              || ||  :|||||||||||||||||||||||| |:: ||:||||::|||:||||| |:
Db        344 PASEISSVLEKGERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRY 403

Qy        993 VVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHH 1051
              :||| :| :   || || |||:|:|::|| |:|||:|||| ||||
Db        404 LVIQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF-------------- 449

Qy       1052 RHRSSSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQS 1106
                | ||                    |:||     | |  : ::  | : |
Db        450 --NSPST-----------------SRTPLLSSLSATSNNSATNCIDRNGG--------- 480

Qy       1107 LPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAA 1166
                |                   |   ||::      | ||||||  : |: ||
Db        481 ---H---------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIY 513

Qy       1167 RPAGAT-LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203
                | : :    |  :|       :   ||:|||||
Db        514 NYISLTAISKLPMDSRYQN-------SHSTAVDNPEYL 544


Search completed: July 22, 2003, 09:10:21
Job time : 32.0157 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-579-593-14
Perfect score: 6841
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6621	96.8	1255	1	ERB2_HUMAN	P04626 homo sapien
2	5833	85.3	1257	1	ERB2_RAT	P06494 rattus norv
3	5817.5	85.0	1254	1	ERB2_MESAU	Q60553 mesocricetu
4	3108	45.4	1210	1	EGFR_HUMAN	P00533 homo sapien
5	3080	45.0	1210	1	EGFR_MOUSE	Q01279 mus musculu
6	2945.5	43.1	1308	1	ERB4_HUMAN	Q15303 homo sapien
7	2929	42.8	1308	1	ERB4_RAT	Q62956 rattus norv
8	2643.5	38.6	1167	1	XMRK_XIPMA	P13388 xiphophorus
9	2387.5	34.9	1342	1	ERB3_HUMAN	P21860 homo sapien
10	2316.5	33.9	1339	1	ERB3_RAT	Q62799 rattus norv
11	1960	28.7	1426	1	EGFR_DROME	P04412 drosophila
12	1708.5	25.0	634	1	ERBB_ALV	P00534 avian leuko
13	1677	24.5	604	1	ERBB_AVIER	P00535 avian eryth
14	1589	23.2	540	1	ERBB_AVIEU	P11273 avian eryth
15	1554	22.7	703	1	EGFR_CHICK	P13387 gallus gall
16	1305	19.1	1323	1	LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.7	245	1	ERB2_MOUSE	P70424 mus musculu
18	730	10.7	1363	1	ILPR_BRALA	O02466 branchiosto
19	710	10.4	1300	1	IRR_MOUSE	Q9wtl4 mus musculu
20	696	10.2	1297	1	IRR_HUMAN	P14616 homo sapien
21	694	10.1	1382	1	INSR_HUMAN	P06213 homo sapien
22	694	10.1	1383	1	INSR_RAT	P15127 rattus norv
23	693	10.1	1477	1	HTK7_HYDAT	Q25197 hydra atten
24	691.5	10.1	1372	1	INSR_MOUSE	P15208 mus musculu
25	691	10.1	1607	1	MIPR_LYMST	Q25410 lymnaea sta
26	688.5	10.1	1300	1	IRR_CAVPO	P14617 cavia porce
27	647	9.5	1367	1	IGIR_HUMAN	P08069 homo sapien
28	629	9.2	1373	1	IGIR_MOUSE	Q60751 mus musculu
29	625.5	9.1	1370	1	IGIR_RAT	P24062 rattus norv
30	620	9.1	1390	1	INSR_AEDAE	Q93105 aedes aegyp
31	614	9.0	2146	1	INSR_DROME	P09208 drosophila
32	597	8.7	987	1	EPB4_HUMAN	P54760 homo sapien
33	594.5	8.7	977	1	EPA2_MOUSE	Q03145 mus musculu
34	591.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
35	588.5	8.6	976	1	EPA2_HUMAN	P29317 homo sapien
36	586.5	8.6	984	1	EPB1_CHICK	Q07494 gallus gall
37	585.5	8.6	984	1	EPB1_HUMAN	P54762 homo sapien
38	583	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae
39	581.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
40	577	8.4	987	1	EPB4_MOUSE	P54761 mus musculu
41	574.5	8.4	1053	1	FAK1_CHICK	Q00944 gallus gall
42	574	8.4	1114	1	RET_HUMAN	P07949 homo sapien
43	569	8.3	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	566	8.3	757	1	HT16_HYDAT	P53356 hydra atten
45	563	8.2	1052	1	FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

```
RESULT 1
ERB2_HUMAN
ID   ERB2_HUMAN     STANDARD;      PRT;  1255 AA.
AC   P04626;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE   surface receptor HER2) (MLN 19).
GN   ERBB2 OR HER2 OR NGL OR NEU.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86118663; PubMed=3003577;
RA   Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA   Saito T., Toyoshima K.;
RT   "Similarity of protein encoded by the human c-erb-B-2 gene to
RT   epidermal growth factor receptor.";
RL   Nature 319:230-234(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86070181; PubMed=2999974;
RA   Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA   McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA   Francke U., Levinson A., Ullrich A.;
RT   "Tyrosine kinase receptor with extensive homology to EGF receptor
RT   shares chromosomal location with neu oncogene.";
RL   Science 230:1132-1139(1985).
RN   [3]
RP   SEQUENCE OF 737-1031 FROM N.A.
RX   MEDLINE=86016729; PubMed=2995967;
RA   Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT   "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT   c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT   human salivary gland adenocarcinoma.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN   [4]
RP   VARIANTS VAL-654 AND VAL-655.
RX   MEDLINE=93194196; PubMed=8095488;
RA   Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT   "Characterization of a new allele of the human ERBB2 gene by allele-
RT   specific competition hybridization.";
RL   Genomics 15:426-429(1993).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
```

```
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
CC       POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC       OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC       ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11767; AAA35808.1; -.
DR   EMBL; M11761; AAA35808.1; JOINED.
DR   EMBL; M11762; AAA35808.1; JOINED.
DR   EMBL; M11763; AAA35808.1; JOINED.
DR   EMBL; M11764; AAA35808.1; JOINED.
DR   EMBL; M11765; AAA35808.1; JOINED.
DR   EMBL; M11766; AAA35808.1; JOINED.
DR   EMBL; M11730; AAA75493.1; -.
DR   EMBL; M12036; AAA35978.1; -.
DR   EMBL; X03363; CAA27060.1; -.
DR   PIR; A25491; A25491.
DR   PIR; A24571; A24571.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3430; ERBB2.
DR   MIM; 164870; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Polymorphism.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1255       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    220    227       BY SIMILARITY.
FT   DISULFID    224    235       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     654    654       I -> V.
FT                                 /FTId=VAR_004077.
FT   VARIANT     655    655       I -> V.
FT                                 /FTId=VAR_004078.
FT   CONFLICT   1170   1170       P -> A (IN REF. 2).
SQ   SEQUENCE   1255 AA;  137909 MW;  39E9DFDA04DCF962 CRC64;

  Query Match             96.8%;  Score 6621;  DB 1;  Length 1255;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1225;  Conservative    7;  Mismatches   19;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |          |   |::     ||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710
```

```
Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

```
RESULT 2
ERB2_RAT
ID   ERB2_RAT       STANDARD;      PRT;  1257 AA.
AC   P06494;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE   receptor-related protein).
GN   ERBB2 OR NEU.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=86118662; PubMed=3945311;
RA   Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT   "The neu oncogene encodes an epidermal growth factor receptor-related
RT   protein.";
RL   Nature 319:226-230(1986).
RN   [2]
RP   SEQUENCE OF 852-905 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   STRUCTURE BY NMR OF 650-668.
RX   MEDLINE=92155181; PubMed=1346763;
RA   Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,
RA   Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT   "Three dimensional structure of the transmembrane region of the proto-
RT   oncogenic and oncogenic forms of the neu protein.";
RL   EMBO J. 11:43-48(1992).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC       THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03362; CAA27059.1; ALT_INIT.
DR   PIR; A24562; TVRTNU.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1257       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    677       POTENTIAL.
FT   DOMAIN      678   1257       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      159    369       CYS-RICH.
FT   DOMAIN      473    646       CYS-RICH.
FT   DOMAIN      722    989       PROTEIN KINASE.
FT   NP_BIND     728    736       ATP (BY SIMILARITY).
FT   BINDING     755    755       ATP (BY SIMILARITY).
FT   ACT_SITE    847    847       BY SIMILARITY.
FT   DISULFID    196    205       BY SIMILARITY.
FT   DISULFID    200    213       BY SIMILARITY.
FT   DISULFID    221    228       BY SIMILARITY.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    237    245       BY SIMILARITY.
FT   DISULFID    241    253       BY SIMILARITY.
FT   DISULFID    256    265       BY SIMILARITY.
FT   DISULFID    269    296       BY SIMILARITY.
FT   DISULFID    300    312       BY SIMILARITY.
FT   DISULFID    316    332       BY SIMILARITY.
FT   DISULFID    335    339       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
```

```
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    578       BY SIMILARITY.
FT   DISULFID    569    586       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    602    625       BY SIMILARITY.
FT   DISULFID    628    636       BY SIMILARITY.
FT   DISULFID    632    644       BY SIMILARITY.
FT   MOD_RES    1141   1141       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1250   1250       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    260    260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     661    661       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1257 AA;  138831 MW;  6129264583011402 CRC64;

  Query Match             85.3%;  Score 5833;  DB 1;  Length 1257;
  Best Local Similarity   85.4%;  Pred. No. 5.4e-305;
  Matches 1082;  Conservative   53;  Mismatches  116;  Indels   16;  Gaps    4;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| |||||||||||||||:|||||||::|||||||||||||||| ||||||||
Db         61 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||||: |||||||| :||||:| |||||
Db        121 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 180

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |   |||||||| ||:| || : ||||| |||| || |:|  ||||||| |||||||||
Db        181 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 240

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||||||||||||||||||||||||||||||||| |||||||||||||| |
Db        241 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 300

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||||||||||::::
Db        301 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 360

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
                || |||||||||||||||||||:|  |||:|||||||||||||||||||||||||||
Db        361 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 420

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              ||||||||::|||||||:|||||||||||||  |||||||||||||||| | |||||||
Db        421 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 480

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||: |||:: ||  || |: ||| ||||||||||||||| |||||
Db        481 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : |||||||||||| | |||| ||||| |||||||   |||
Db        541 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 600

Qy        599 RCPFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRA 654
              |||                |::     ||||| ||||||||||||||::|||||||||
Db        601 RCPSG----------VKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRA 650

Qy        655 SPLTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPN 714
              ||:| |:: |||:|| ::| || ||||||:|||||||||||||||||||||||||||||
Db        651 SPVTFIIATVVGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPN 710

Qy        715 QAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIL 774
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIL 770

Qy        775 DEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQ 834
              |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||:|
Db        771 DEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQ 830

Qy        835 IAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKW 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 IAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKW 890

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 950

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
              ||||||||||||||||||||||||||||||||||||||||||||||:||:|||||||||
Db        951 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSL 1010

Qy       1015 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1074
              |||||||||||||||||||||| ||| || |   | ||||||||||:||||||||||
Db       1011 LEDDDMGDLVDAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEE 1070

Qy       1075 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1134
                |||||||||||||||||| || ||||||  || ||||||||||:||| ||||||
Db       1071 GPPRSPLAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYV 1130

Qy       1135 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1194
              ||| |||||||||| :|:|||| |||||  |||||||| ||||||||||||||||||
Db       1131 APLACSPQPEYVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFG 1190

Qy       1195 GAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254
              |||||||| |: | |  ||| |||||||||||||:  |:| ||| |:|||||||||
Db       1191 GAVENPEYLVPREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEY 1250

Qy       1255 LGLDVPV 1261
              |||||||
Db       1251 LGLDVPV 1257


RESULT 3
ERB2_MESAU
ID   ERB2_MESAU     STANDARD;      PRT;  1254 AA.
AC   Q60553;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN   ERBB2 OR NEU.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Nerve;
RX   MEDLINE=94193007; PubMed=7908275;
RA   Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA   Yamazaki Y., Ishikawa T.;
RT   "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL   Gene 140:251-255(1994).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
```

```
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D16295; BAA03801.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1254       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1254       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      158    368       CYS-RICH.
FT   DOMAIN      472    644       CYS-RICH.
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1247   1247       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     658    658       V -> E (IN ONCOGENIC NEU).
FT   VARIANT     659    659       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1254 AA;  138252 MW;  974C3791C21F2BE1 CRC64;
```

```
  Query Match             85.0%;  Score 5817.5;  DB 1;  Length 1254;
  Best Local Similarity   85.0%;  Pred. No. 3.6e-304;
  Matches 1075;  Conservative   62;  Mismatches  113;  Indels   15;  Gaps    3;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| | |||||||| |||: ||||||||||||||||||||::|||||||||||||||
Db          1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:||||:||| |||||||||||||||||| |||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||:| |  || :| |||||||||||||||||||||: |||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||| ||:| || : ||| | ||||:|| |:        :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||:|||||||||:|||||||||||||||||| ||||||||||:::::
Db        301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||:|:|   ||| |||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||:|||||:||:||||| |||||| |||||||||||||| ||| ||||||||||
Db        421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||: |  |:||  :  ||: ||| ||||||||||||||| |||||||
Db        481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |:|||| :|||||||| :||||||||||||| : || | |||| || |||| |||||||
Db        541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |              |   |::     ||||| ||||||||||||||::||||||||||
Db        601 PSG----------VKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              |||:: |||||| :|:||| |||||||:|||||||||||||||||||||||||||||||
Db        651 ATSIIATVVGILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||:|||||||||||||||||||||||||||||||||||:|||||||||||||:|||
Db        771 AYVMAGLGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLE 1010
```

```
Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||| ||||||||    | ||||||||||||:||||:||| ||
Db       1011 DDDMGDLVDAEEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEP 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||:|:||||| || ||:    | ||||||||||:|||:||||||||
Db       1071 PRSPLAPSEGAGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              | ||||||||||:|||||   |||||  |||||||| |||||||||||||| ||||
Db       1131 LACSPQPEYVNQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||| |:||:| ||| |||  |||||||||||| |||:||:||:||||||||||||
Db       1191 VENPEYLVPRGGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLG 1249

Qy       1257 LDVPV 1261
              |||||
Db       1250 LDVPV 1254
```

```
RESULT 4
EGFR_HUMAN
ID   EGFR_HUMAN     STANDARD;      PRT;  1210 AA.
AC   P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC   O00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE   protein-tyrosine kinase ErbB-1).
GN   EGFR OR ERBB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=84219729; PubMed=6328312;
RA   Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA   Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA   Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT   "Human epidermal growth factor receptor cDNA sequence and aberrant
RT   expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL   Nature 309:418-425(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=95382957; PubMed=7654368;
RA   Ilekis J.V., Stark B.C., Scoccia B.;
RT   "Possible role of variant RNA transcripts in the regulation of
RT   epidermal growth factor receptor expression in human placenta.";
RL   Mol. Reprod. Dev. 41:149-156(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97078686; PubMed=8918811;
RA   Reiter J.L., Maihle N.J.;
RT   "A 1.8 kb alternative transcript from the human epidermal growth
RT   factor receptor gene encodes a truncated form of the receptor.";
RL   Nucleic Acids Res. 24:4050-4056(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97256547; PubMed=9103388;
RA   Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT   "Expression of a truncated epidermal growth factor receptor-like
RT   protein (TEGFR) in ovarian cancer.";
RL   Gynecol. Oncol. 65:36-41(1997).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC   TISSUE=Placenta;
RX   MEDLINE=21100872; PubMed=11161793;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative EGFR transcripts encoding truncated receptor
RT   isoforms.";
RL   Genomics 71:1-20(2001).
RN   [6]
RP   SEQUENCE OF 575-687 FROM N.A.
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Human and mouse alternative EGFR transcripts encoding only the
RT   extracellular domain of the receptor.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 713-924 FROM N.A.
RX   MEDLINE=84196372; PubMed=6326261;
RA   Lin C.R., Chen W.S., Kruiger W., Stolarsky L.S., Weber W.,
RA   Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT   "Expression cloning of human EGF receptor complementary DNA: gene
RT   amplification and three related messenger RNA products in A431
RT   cells.";
RL   Science 224:843-848(1984).
RN   [8]
RP   SEQUENCE OF 150-962 FROM N.A.
RX   MEDLINE=84245835; PubMed=6330563;
RA   Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA   Roe B.A., Merlino G.T., Pastan I.;
RT   "Human epidermal growth factor receptor cDNA is homologous to a
RT   variety of RNAs overproduced in A431 carcinoma cells.";
RL   Nature 309:806-810(1984).
RN   [9]
RP   SEQUENCE OF 1028-1210 FROM N.A.
RX   MEDLINE=85046483; PubMed=6093780;
RA   Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA   O'Malley B.W.;
RT   "Isolation of an evolutionarily conserved epidermal growth factor
RT   receptor cDNA from human A431 carcinoma cells.";
RL   Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN   [10]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=88217333; PubMed=3329716;
RA   Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA   Waterfield M.D.;
RT   "The human EGF receptor gene: structure of the 110 kb locus and
RT   identification of sequences regulating its transcription.";
RL   Oncogene Res. 1:375-396(1987).
RN   [11]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=91107677; PubMed=1988448;
RA   Haley J.D., Waterfield M.D.;
RT   "Contributory effects of de novo transcription and premature
RT   transcript termination in the regulation of human epidermal growth
RT   factor receptor proto-oncogene RNA synthesis.";
RL   J. Biol. Chem. 266:1746-1753(1991).
RN   [12]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=85270438; PubMed=2991899;
RA   Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT   "Characterization and sequence of the promoter region of the human
RT   epidermal growth factor receptor gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN   [13]
RP   SEQUENCE OF 540.
RA   Kohda D.;
RL   Submitted (SEP-1997) to the SWISS-PROT data bank.
RN   [14]
RP   RECEPTOR ACTIVITY.
RX   MEDLINE=84191554; PubMed=6325948;
RA   Mroczkowski B., Mosig G., Cohen S.;
RT   "ATP-stimulated interaction between epidermal growth factor receptor
```

```
RT   and supercoiled DNA.";
RL   Nature 309:270-273(1984).
RN   [15]
RP   PHOSPHORYLATION.
RX   MEDLINE=89278137; PubMed=2543678;
RA   Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA   Howk R., Givol D., Ullrich A., Schlessinger J.;
RT   "All autophosphorylation sites of epidermal growth factor (EGF)
RT   receptor and HER2/neu are located in their carboxyl-terminal tails.
RT   Identification of a novel site in EGF receptor.";
RL   J. Biol. Chem. 264:10667-10671(1989).
RN   [16]
RP   CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
RP   ASN-528.
RX   MEDLINE=96398132; PubMed=8962717;
RA   Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
RT   "Analysis of the glycosylation patterns of the extracellular domain of
RT   the epidermal growth factor receptor expressed in Chinese hamster
RT   ovary fibroblasts.";
RL   Growth Factors 13:121-132(1996).
RN   [17]
RP   CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
RP   ASN-603.
RX   MEDLINE=20198209; PubMed=10731668;
RA   Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
RT   "Characterization of the N-oligosaccharides attached to the atypical
RT   Asn-X-Cys sequence of recombinant human epidermal growth factor
RT   receptor.";
RL   J. Biochem. 127:65-72(2000).
RN   [18]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=98225196; PubMed=9556602;
RA   Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
RT   "Disulfide bond structure of human epidermal growth factor receptor.";
RL   J. Biol. Chem. 273:11150-11157(1998).
RN   [19]
RP   REVIEW.
RX   MEDLINE=87297456; PubMed=3039909;
RA   Carpenter G.;
RT   "Receptors for epidermal growth factor and other polypeptide
RT   mitogens.";
RL   Annu. Rev. Biochem. 56:881-914(1987).
CC   -!- FUNCTION: Receptor for EGF, but also for other members of the EGF
CC       family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
CC       EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
CC       involved in the control of cell growth and differentiation.
CC   -!- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
CC       secreted.
CC   -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
CC       truncated isoform/TEGFR, 3/p110 and 4; are produced by
CC       alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
CC       expressed in ovarian cancers.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00588; CAA25240.1; -.
DR   EMBL; U95089; AAB53063.1; -.
DR   EMBL; U48722; AAC50802.1; -.
DR   EMBL; U48723; AAC50804.1; -.
DR   EMBL; U48724; AAC50796.1; -.
DR   EMBL; U48725; AAC50797.1; -.
DR   EMBL; U48726; AAC50798.1; -.

  Query Match             45.4%;  Score 3108;  DB 1;  Length 1210;
  Best Local Similarity   49.6%;  Pred. No. 4e-159;
  Matches  633;  Conservative  169;  Mismatches  351;  Indels  124;  Gaps   26;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
              || || |   |    :|| ||  ||    : | |   |: ::  |:|| ||||:||:  |
Db         14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
                ||||: |||| |||||| | | ::||: |:|:||   :|::|||||| | |
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
                |:  ||:|| :|:| ||| | |    || ||  ::| |:||   :    :::   |
Db        127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy        189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
                :|  | | |    |||   |:|| ||: :||  |: ||:|  |:|||| |||||||||
Db        184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy        248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
              : |||| |  |      |:  || |: ||  |::   ||||:|:|||:||  || ||: ||
Db        244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy        308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
              |||   |   : |:  ||| ::|:||  || :|| |:|:   | :|  |  | :: |
Db        304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy        367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
              |  | | |  || :| ||  ::| || |::| : :|::|||||:| | |||::  ||  |:
Db        362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy        427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
              ||::||||   :| :|| :  | |: ||||||:|:  | :|  | :||: :|: | :||
Db        422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy        487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
                    :|| |: |   |  || ||:   |||| |  ||:|   ||:|||::| :
Db        482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNL 541

Qy        547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPF---- 602
              |:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||  ||
Db        542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy        603 -NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIV 661
               ||  | |       | |:       |  |  |||: |     :|||       | ||
Db        602 ENNTLV--W---KYADAGHV-------CHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIA 647

Qy        662 SAVVG---ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQM 718
              : :||   :|||| ||:    |  ||: :|| |:|||||| |||||||||  |||| :
Db        648 TGMVGALLLLLVVALGIG---LFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALL 704

Qy        719 RILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAY 778
              ||||||| :|:|||||||||||||:|||:|| |||||||| ||| ||||||||||||||
Db        705 RILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAY 764

Qy        779 VMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKG 838
              ||| | :|:| ||||||||||||:|||||:|||||:|||::  :||| |||||:|||||
Db        765 VMASVDNPHVCRLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKG 824

Qy        839 MSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALE 898
              |:|||| |||||||||||||||:| ||||||||||:||  :| |||||||||||||||||
Db        825 MNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALE 884

Qy        899 SILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVY 958
              ||| | :|||||||||||||||||||:|||||||| ||  :|||||||||||||||||||
Db        885 SILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVY 944
```

```
Qy        959 MIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLED 1017
              ||||||||||:: ||:||||: |||:||||||::||| :| :   || || |||:|:::
Db        945 MIMVKCWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDE 1004

Qy       1018 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAP 1077
              :|| |:|||:|||:|||||                                        |
Db       1005 EDMDDVVDADEYLIPQQGFF---------------------------------SSPSTS 1030

Qy       1078 RSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVA 1135
              |:||  |  | |:  :  :      |||| |  : | ||||| |||  |  ::  | ::
Db       1031 RTPLLSSLSATSN--NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL- 1087

Qy       1136 PLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGG 1195
                    | |||:|| |  :|    : |:  :|          :| ::   :|
Db       1088 -----PVPEYINQ-SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHST 1131

Qy       1196 AVENPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGA 1238
              || ||||| | |          |   :  ||:  :| |         :|       |:
Db       1132 AVGNPEYLNTVQ---------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAK 1182

Qy       1239 PPSTFKGTPTAENPEYL 1255
              |   |||: |||| |||
Db       1183 PNGIFKGS-TAENAEYL 1198
```

```
RESULT 5
EGFR_MOUSE
ID   EGFR_MOUSE     STANDARD;      PRT;  1210 AA.
AC   Q01279;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=93026370; PubMed=1408137;
RA   Avivi A., Skorecki K., Yayon A., Givol D.;
RT   "Promoter region of the murine fibroblast growth factor receptor 2
RT   (bek/KGFR) gene.";
RL   Oncogene 7:1957-1962(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX   MEDLINE=93126380; PubMed=7678348;
RA   Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT   "Expression of the epidermal growth factor receptor gene is regulated
RT   in mouse blastocysts during delayed implantation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RA   Hibbs M.L.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/C3; TISSUE=Liver;
RX   MEDLINE=94170986; PubMed=8125255;
RA   Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA   Jenkins N.A., Lee D.C.;
RT   "The mouse waved-2 phenotype results from a point mutation in the EGF
RT   receptor tyrosine kinase.";
RL   Genes Dev. 8:399-413(1994).
RN   [5]
RP   SEQUENCE OF 1-714 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=91232866; PubMed=2030916;
RA   Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT   "Comparison of EGF receptor sequences as a guide to study the ligand
RT   binding site.";
RL   Oncogene 6:673-676(1991).
RN   [6]
RP   SEQUENCE OF 969-1117 FROM N.A.
RC   STRAIN=C3H;
RA   Eisinger D.P., Serrero G.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X78987; CAA55587.1; -.
DR   EMBL; U03425; AAA17899.1; -.
DR   EMBL; X59698; CAA42219.1; -.
DR   EMBL; L06864; AAA53029.1; -.
DR   EMBL; Z12608; CAA78249.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1210       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       25    647       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    648    670       POTENTIAL.
FT   DOMAIN      671   1210       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       75    300       APPROXIMATE.
FT   REPEAT      390    600       APPROXIMATE.
FT   DOMAIN     1028   1071       SER-RICH.
FT   DOMAIN      714    981       PROTEIN KINASE.
FT   NP_BIND     720    728       ATP (BY SIMILARITY).
FT   BINDING     747    747       ATP (BY SIMILARITY).
FT   ACT_SITE    839    839       BY SIMILARITY.
FT   DISULFID    190    199       BY SIMILARITY.
FT   DISULFID    194    207       BY SIMILARITY.
FT   DISULFID    215    223       BY SIMILARITY.
FT   DISULFID    219    231       BY SIMILARITY.
FT   DISULFID    232    240       BY SIMILARITY.
FT   DISULFID    236    248       BY SIMILARITY.
FT   DISULFID    251    260       BY SIMILARITY.
FT   DISULFID    264    291       BY SIMILARITY.
FT   DISULFID    295    307       BY SIMILARITY.
FT   DISULFID    311    326       BY SIMILARITY.
FT   DISULFID    329    333       BY SIMILARITY.
```

```
FT   DISULFID    506    515       BY SIMILARITY.
FT   DISULFID    510    523       BY SIMILARITY.
FT   DISULFID    526    535       BY SIMILARITY.
FT   DISULFID    539    555       BY SIMILARITY.
FT   DISULFID    558    571       BY SIMILARITY.
FT   DISULFID    562    579       BY SIMILARITY.
FT   DISULFID    582    591       BY SIMILARITY.
FT   DISULFID    595    617       BY SIMILARITY.
FT   DISULFID    620    628       BY SIMILARITY.
FT   DISULFID    624    636       BY SIMILARITY.
FT   MOD_RES     680    680       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   MOD_RES    1092   1092       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1110   1110       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1172   1172       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1197   1197       PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT                                (BY SIMILARITY).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    175    175       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    352    352       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    444    444       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     19     19       C -> S (IN REF. 2).
FT   CONFLICT    539    539       C -> W (IN REF. 5).
FT   CONFLICT    991    991       L -> F (IN REF. 4).
FT   CONFLICT   1116   1117       HP -> DR (IN REF. 6).
SQ   SEQUENCE   1210 AA;  134853 MW;  690E20D46DF2D2F5 CRC64;

  Query Match             45.0%;  Score 3080;  DB 1;  Length 1210;
  Best Local Similarity   49.5%;  Pred. No. 1.3e-157;
  Matches  635;  Conservative  161;  Mismatches  360;  Indels  128;  Gaps   27;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPF 602
```

```
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 -----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPL 657
Db        598 GIMGENNTLV--WKYA----------DANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKI 645

Qy        658 TSIVSAVVGILLVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
Db        646 PSIATGIVGGLLFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQA 704

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
Db        705 HLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDE 764

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
Db        765 AYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIA 824

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
Db        825 KGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMA 884

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
Db        885 LESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTID 944

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLL 1015
Db        945 VYMIMVKCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALM 1004

Qy       1016 EDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEE 1075
Db       1005 DEEDMEDVVDADEYLIPQQGFF----------------NSPST----------------- 1031

Qy       1076 APRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGY 1133
Db       1032 -SRTPLLSSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAF 1086

Qy       1134 VAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAF 1193
Db       1087 L------PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PH 1129

Qy       1194 GGAVENPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PER 1236
Db       1130 SNAVGNPEYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKE 1180

Qy       1237 GAPPSTFKGTPTAENPEYLGLDVP 1260
Db       1181 TKPNGIFKG-PTAENAEYLRVAPP 1203
```

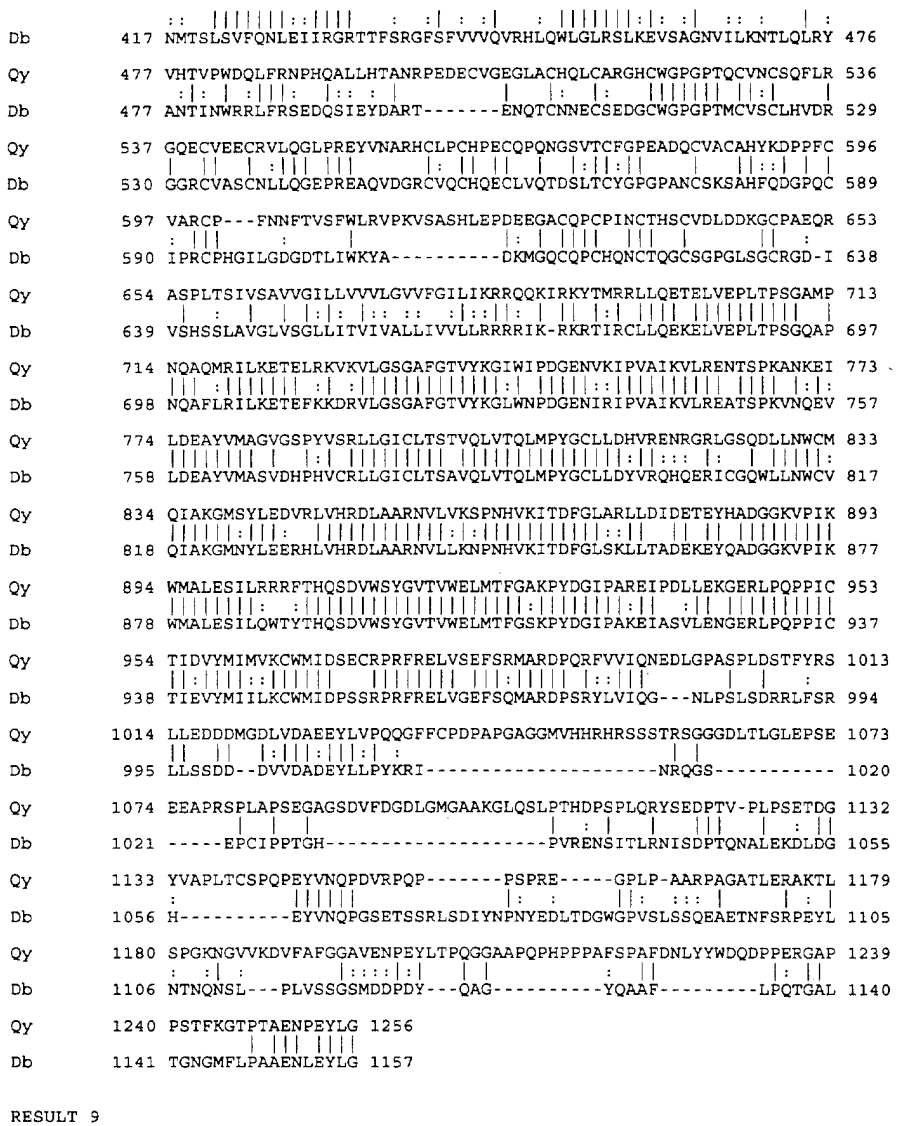
```
RESULT 6
ERB4_HUMAN
ID   ERB4_HUMAN     STANDARD;      PRT;  1308 AA.
AC   Q15303;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE   (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN   ERBB4 OR HER4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM JM-A).
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=93189574; PubMed=8383326;
RA   Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA   Foy L., Neubauer M.G., Shoyab M.;
RT   "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
```

```
RT   epidermal growth factor receptor family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97476287; PubMed=9334263;
RA   Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA   Klagsbrun M.;
RT   "A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
RT   tissue distribution and differential processing in response to
RT   phorbol ester.";
RL   J. Biol. Chem. 272:26761-26768(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC       FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC       BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC       NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC       CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC       KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC       PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC       LUNG, SALIVARY GLAND, AND PANCREAS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07868; AAB59446.1; -.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3432; ERBB4.
DR   MIM; 600543; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN      676   1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      186    334       CYS-RICH.
FT   DOMAIN      496    633       CYS-RICH.
FT   DOMAIN      718    985       PROTEIN KINASE.
FT   NP_BIND     724    732       ATP (BY SIMILARITY).
FT   BINDING     751    751       ATP (BY SIMILARITY).
FT   ACT_SITE    843    843       BY SIMILARITY.
FT   DISULFID    189    197       BY SIMILARITY.
FT   DISULFID    193    205       BY SIMILARITY.
FT   DISULFID    213    221       BY SIMILARITY.
FT   DISULFID    217    229       BY SIMILARITY.
FT   DISULFID    230    238       BY SIMILARITY.
FT   DISULFID    234    246       BY SIMILARITY.
FT   DISULFID    249    258       BY SIMILARITY.
FT   DISULFID    262    289       BY SIMILARITY.
FT   DISULFID    293    304       BY SIMILARITY.
FT   DISULFID    308    323       BY SIMILARITY.
FT   DISULFID    326    330       BY SIMILARITY.
FT   DISULFID    503    512       BY SIMILARITY.
FT   DISULFID    507    520       BY SIMILARITY.
FT   DISULFID    523    532       BY SIMILARITY.
FT   DISULFID    536    552       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    559    577       BY SIMILARITY.
FT   DISULFID    580    589       BY SIMILARITY.
FT   DISULFID    593    614       BY SIMILARITY.
FT   DISULFID    617    625       BY SIMILARITY.
FT   DISULFID    621    633       BY SIMILARITY.
FT   MOD_RES    1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    626    648       NGPTSHDCIYYPWTGHSTLPQHA -> IGSSIEDCIGLMD
FT                                 (IN ISOFORM JM-B).
SQ   SEQUENCE   1308 AA;  146807 MW;  5E4AE80985D88761 CRC64;

  Query Match           43.1%;  Score 2945.5;  DB 1;  Length 1308;
  Best Local Similarity  44.9%;  Pred. No. 2.2e-150;
  Matches  609;  Conservative  181;  Mismatches  378;  Indels  189;  Gaps   30;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
               :  |  ||||: ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
               |           ||:|| |::||||| ||| : :|  ||| ||| |:||        |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
               || |  |  |  | | |||| :   ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |  ||    ||  ||:  |  :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||   ||  |:| : |: |:  | : | |:|   : :          :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :    : ||:| || |: |||:| | :|| :: | |
```

```
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483

Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543

Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCP- 601

Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        602 ----FNNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGC--------- 648

Db        596 GLQGANSFIFKY-----------ADPDRE--CHPCHPNCTQGCNGPTSHDCIYYPWTGHS 642

Qy        649 --PAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEP 705

Db        643 TLPQHAR-TPL--IAAGVIGGLFILVIVGLTFAVYVRRKSIK-KKRALRRFL-ETELVEP 697

Qy        706 LTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENT 765

Db        698 LTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETT 757

Qy        766 SPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGS 825

Db        758 GPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGS 817

Qy        826 QDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHA 885

Db        818 QLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNA 877

Qy        886 DGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGE 945

Db        878 DGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGE 937

Qy        946 RLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPAS 1004

Db        938 RLPQPPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPS 997

Qy       1005 PLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGD 1064

Db        998 PNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS---- 1046

Qy       1065 LTLGLEPSEEEAPRS---------------------------PLAP-SEGAGSDVFDGD 1095

Db       1047 -EIGHSPPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDS 1105

Qy       1096 LGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQ 1148

Db       1106 CCNGTLRKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNP 1165

Qy       1149 PDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGG 1208

Db       1166 VE---------ENPFVSRR---------------KNGDLQ-------ALDNPEYHNASNG 1194

Qy       1209 AAPQPHPPPA---------------------------------FSPAFDNLYYWDQDPP 1234

Db       1195 ------PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLP 1248

Qy       1235 ERGA--PPSTFKGTPT-------------AENPEYL 1255

Db       1249 PRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 7
ERB4_RAT
ID   ERB4_RAT       STANDARD;      PRT;  1308 AA.
AC   Q62956; Q9Z2N7;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN   ERBB4 OR TYRO-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=98221155; PubMed=9553078;
RA   Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA   Marchionni M.A., Kelly R.A.;
RT   "Neuregulins promote survival and growth of cardiac myocytes.
RT   Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT   ventricular myocytes.";
RL   J. Biol. Chem. 273:10261-10269(1998).
RN   [2]
RP   SEQUENCE OF 848-901 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   SEQUENCE OF 1031-1198 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC       NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC       OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC       RETICULAR NUCLEUS OF THE THALLAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC       HEART.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF041838; AAD08899.1; -.
DR   EMBL; U52531; AAC53051.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
```

```
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN      676   1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      186    334       CYS-RICH.
FT   DOMAIN      496    633       CYS-RICH.
FT   DOMAIN      718    985       PROTEIN KINASE.
FT   NP_BIND     724    732       ATP (BY SIMILARITY).
FT   BINDING     751    751       ATP (BY SIMILARITY).
FT   ACT_SITE    843    843       BY SIMILARITY.
FT   DISULFID    189    197       BY SIMILARITY.
FT   DISULFID    193    205       BY SIMILARITY.
FT   DISULFID    213    221       BY SIMILARITY.
FT   DISULFID    217    229       BY SIMILARITY.
FT   DISULFID    230    238       BY SIMILARITY.
FT   DISULFID    234    246       BY SIMILARITY.
FT   DISULFID    249    258       BY SIMILARITY.
FT   DISULFID    262    289       BY SIMILARITY.
FT   DISULFID    293    304       BY SIMILARITY.
FT   DISULFID    308    323       BY SIMILARITY.
FT   DISULFID    326    330       BY SIMILARITY.
FT   DISULFID    503    512       BY SIMILARITY.
FT   DISULFID    507    520       BY SIMILARITY.
FT   DISULFID    523    532       BY SIMILARITY.
FT   DISULFID    536    552       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    559    577       BY SIMILARITY.
FT   DISULFID    580    589       BY SIMILARITY.
FT   DISULFID    593    614       BY SIMILARITY.
FT   DISULFID    617    625       BY SIMILARITY.
FT   DISULFID    621    633       BY SIMILARITY.
FT   MOD_RES    1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1062   1062       S -> N (IN REF. 3).
FT   CONFLICT   1080   1082       PYT -> SYR (IN REF. 3).
SQ   SEQUENCE   1308 AA;  146957 MW;  D944BB0996A08B41 CRC64;

  Query Match             42.8%;  Score 2929;  DB 1;  Length 1308;
  Best Local Similarity   44.6%;  Pred. No. 1.7e-149;
  Matches  606;  Conservative  188;  Mismatches  388;  Indels  176;  Gaps   29;

Qy          1 MELA-ALCRWGLLL--ALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQ 57
              |:||  |  || ||     :||  || ||: ||   :  |      ||  |: |:||
Db          1 MKLATGLWVWGSLLVAARTVQPSASQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVM 60

Qy         58 GNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVL 117
              ||||:|  :  |  ||||: |:|| ||||:| || | :||: |||:|||:|:|| ||||:
Db         61 GNLEITSIEHNRDLSFLRSIREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIF 120

Qy        118 DNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNN 177
               |     |          ||:|| |::|||| ||| : :|  ||| ||| |:||
Db        121 LNYRKDGNF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPW 171

Qy        178 QLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCC 236
              :||: |  |  |  |  | |||| :  ||:||||||  |  || ||  :|||
Db        172 PSNMTLVSTIGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCC 230

Qy        237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
              | :|| ||:||| :|| ||::|| || |   ||     ||  ||:   |   :||:|| ||
Db        231 HRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCV 290

Qy        297 TACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF 356
                ||:|::  |  ||   ||    ||  |:| : |: |:  | : | |:|    :
Db        291 KKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTV 348

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
                    :|  |  || |:| ||     |||  :    : ||:| || |: ||||:| |  ||
Db        349 DSSNIDKFINCTKINGNLIFLVTGIHGDPYNAIDAIDPEKLNVFRTVREITGFLNIQTWP 408

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
              :: | ||| ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||:
Db        409 PNMTDFSVPSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCY 468

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
              ||: |  ||   :| ::   ||  :  ||: |: ||:   |||||  ||::| :| |
Db        469 YHTINWTTLFSTVNQRIVIRDNRRAENCTAEGMVCNHLCSNDGCWGPGPDQCLSCRRFSR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPF 595
              |: |:| | :  |  ||: |  |: |  :|:  ::| :|| ||  | |  |:|:|| |
Db        529 GKICIESCNLYDGEFREFENGSICVECDSQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPN 588

Qy        596 CVARCP-----FNNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGC-- 648
              || :||      |:|  :               |::  | ||  |||  |      |
Db        589 CVEKCPDVLQGANSFIFKY------------ADQDRECHPCHPNCTQGCNGPTSHDCIY 635

Qy        649 ---------PAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQ 698
                      |  | :||  | : |: |: ::|:: : | : ::|:  | :|  :|| |
Db        636 YPWTGHSTLPQHAR-TPL--IAAGVIGGLFILVIMALTFAVYVRRKSIK-KKRALRRFL- 690

Qy        699 ETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAI 758
              ||||||||||||  ||||:|||||||::|||||||||||||||||:|:|| |||||||
Db        691 ETELVEPLTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAI 750

Qy        759 KVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRE 818
              |:| | | |||| | :||| :|| |  |:: ||||:||: |:|||||||:||||::| |
Db        751 KILNETTGPKANVEFMDEALIMASVDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHE 810

Qy        819 NRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDI 878
              ::  :||| |||||:|||||| ||| ||||||||||||||||||||||||||||||||:
Db        811 HKDNIGSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEG 870

Qy        879 DETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIP 938
              || ||:|||||:||||||| |  |:|||||||||||||:||||||| |||||| ||||
Db        871 DEKEYNADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIP 930

Qy        939 DLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNE 998
              |||||||||||||||||||::|||||||:: ||:|:|| :|||||||||||::||| :
Db        931 DLLEKGERLPQPPICTIDVYIVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGD 990

Qy        999 D-LGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDP--------------- 1041
              | :   || || |:::||:::|: |::|||||||| | | | |
Db        991 DRMKLPSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTSRTRIDSNRSEIG 1049

Qy       1042 -------APGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDG 1094
                    | :|    ::    :|: |   : :    :    | :|:|  :|| :::||
Db       1050 HSPPPAYTPMSGSQFVYQDGGFATQQG---MPMPYTATTSTIPEAPVA--QGATAEMFDD 1104

Qy       1095 DLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVN 1147
                |  :     : |  |||| |||| |         : :||: |:   |: ||:|
Db       1105 SCCNGTLRKPVVPHVQEDSSTQRYSADPTVFAPERNPRAELDEEGYMTPMHDKPKQEYLN 1164

Qy       1148 QPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQG 1207
                :          | |  : |                  ||| ::       |::|||| :
Db       1165 PVE---------ENPFVSRR--------------KNGDLQ-------ALDNPEYHSASS 1193

Qy       1208 GAAPQPHPPPA---------------------------------FSPAFDNLYYWDQDP 1233
```

```
                |      || |                                        ||||  ||:
Db       1194 G------PPKAEDEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPDYWNHSL 1247

Qy       1234 PERGA--PPSTFKGTPT-------------AENPEYL 1255
              | |     |   :   |               |||||||
Db       1248 PPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 8
XMRK_XIPMA
ID   XMRK_XIPMA     STANDARD;      PRT;  1167 AA.
AC   P13388;
DT   01-JAN-1990 (Rel. 13, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN   XMRK OR TU.
OS   Xiphophorus maculatus (Southern platyfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8083;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90015140; PubMed=2797166;
RA   Wittbrodt J., Adam D., Malitschek B., Maueler W., Raulf F.,
RA   Telling A., Robertson S.M., Schartl M.;
RT   "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT   inducing Tu locus in Xiphophorus.";
RL   Nature 341:415-421(1989).
RN   [2]
RP   REVISION TO 515.
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16891; CAA34770.2; -.
DR   PIR; S06142; S06142.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT   SIGNAL        1     25
FT   CHAIN        26   1167       MELANOMA RECEPTOR PROTEIN-TYROSINE
FT                                KINASE.
FT   DOMAIN        26    642       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     643    665       POTENTIAL.
FT   DOMAIN       666   1167       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       710    977       PROTEIN KINASE.
FT   NP_BIND      716    724       ATP (BY SIMILARITY).
FT   BINDING      743    743       ATP (BY SIMILARITY).
FT   ACT_SITE     835    835       BY SIMILARITY.
FT   DISULFID     195    204       BY SIMILARITY.
FT   DISULFID     199    212       BY SIMILARITY.
FT   DISULFID     220    228       BY SIMILARITY.
FT   DISULFID     224    236       BY SIMILARITY.
FT   DISULFID     237    245       BY SIMILARITY.
FT   DISULFID     241    253       BY SIMILARITY.
FT   DISULFID     256    265       BY SIMILARITY.
FT   DISULFID     269    296       BY SIMILARITY.
FT   DISULFID     300    311       BY SIMILARITY.
FT   DISULFID     315    330       BY SIMILARITY.
FT   DISULFID     333    337       BY SIMILARITY.
FT   DISULFID     504    513       BY SIMILARITY.
FT   DISULFID     508    521       BY SIMILARITY.
FT   DISULFID     524    533       BY SIMILARITY.
FT   DISULFID     537    553       BY SIMILARITY.
FT   DISULFID     556    569       BY SIMILARITY.
FT   DISULFID     560    577       BY SIMILARITY.
FT   DISULFID     593    615       BY SIMILARITY.
FT   DISULFID     618    626       BY SIMILARITY.
FT   DISULFID     622    634       BY SIMILARITY.
FT   CARBOHYD     114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1167 AA;  129934 MW;  4793E0749DC1D55A CRC64;

  Query Match             38.6%;  Score 2643.5;  DB 1;  Length 1167;
  Best Local Similarity   44.6%;  Pred. No. 2.8e-134;
  Matches  569;  Conservative  165;  Mismatches  392;  Indels  151;  Gaps   30;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              ||| :  ||| |      ||    :|| ||  :: :    : |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||   |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: | |: |
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 | ::|       ||::||| :||||| ||| :  || ||  :|| | ||  | :
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
              : ||    | |  |  |    ||    ||  |: :||  |  ||:|| | |||:|
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              || |||||: :|||||  ||  | |:  ||    |:  : : : ||  :|||||:||
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
              || ||: |: |:|   |      ||  |:| : |: |   | :|| |:|:  : :|:  :
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
                |  :  |: | || | :     ||:|||      : || |    |::|||||| |  ||:
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
```



RESULT 9
ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Iesing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC ---

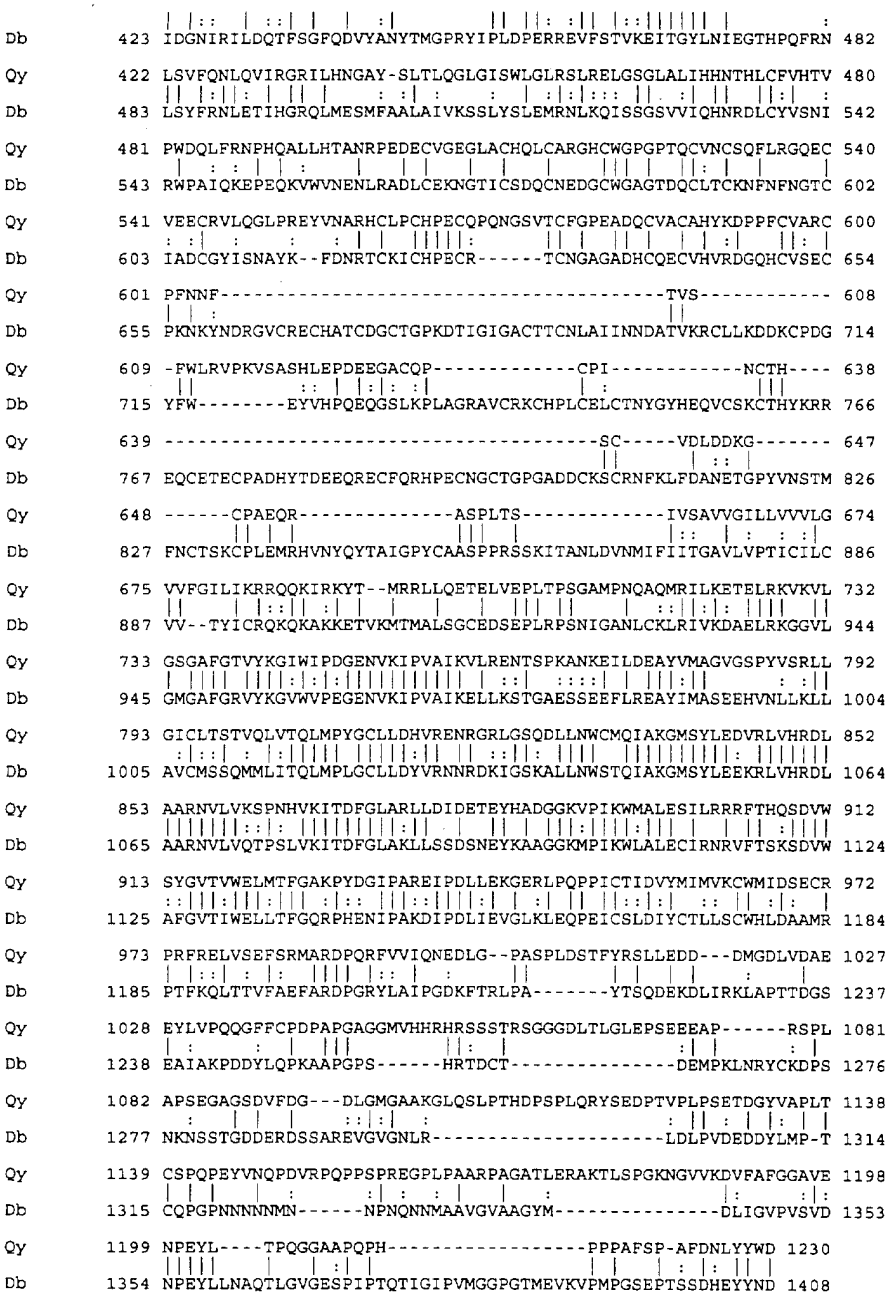
CC ---
DR EMBL; M29366; AAA35790.1; -.
DR EMBL; M34309; AAA35979.1; -.
DR EMBL; S61953; AAB26935.1; -.
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

```
KW   Alternative splicing.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1342       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN       20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    644    664       POTENTIAL.
FT   DOMAIN      665   1342       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      709    966       PROTEIN KINASE.
FT   NP_BIND     715    723       ATP (BY SIMILARITY).
FT   BINDING     742    742       ATP (BY SIMILARITY).
FT   ACT_SITE    834    834       BY SIMILARITY.
FT   DISULFID    186    194       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    552    565       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    141    183       EILSGGVYIEKNDKLCHMDTIDWRDIVRDRDAEIVVKDNGR
FT                                SC -> GQFPMVPSGLTPQPAQDWYLLDDDPRLLTLSASSK
FT                                VPVTLAAV (IN SHORT ISOFORM).
FT   VARSPLIC    184   1342       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    560    560       E -> G (IN REF. 2).
FT   CONFLICT   1064   1064       E -> G (IN REF. 2).
SQ   SEQUENCE   1342 AA;  148097 MW;  7201E7F66CA374BD CRC64;

  Query Match             34.9%;  Score 2387.5;  DB 1;  Length 1342;
  Best Local Similarity   39.9%;  Pred. No. 1.7e-120;
  Matches  527;  Conservative  193;  Mismatches  459;  Indels  141;  Gaps   34;

Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
              ||| :|         |  || ||   | :    |      |  ||: |:|| ||||:
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
              || ||||| |:|| |||:| |:   :||  ||:|||||::  :|: |:     || |
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
                 |   ||:|:|  ||||| ||| |::| :||: ||| |:||     :     ::  :
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
              |:| ||  :||| |||  |||||:||:|:||   |  | || |  |||::|| ||:|
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
              |: :|| || ||| || |   ||  : ||  ||:  |||  :| :|  || :||:|::
Db        238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFQLEPNPHTKYQYGGVCVASCPHNFV-V 296

Qy        307 DVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF--IGITELE- 363
              |  ||   ||    || ::| : || |   | : | | |     : |:|  :  : ::
Db        297 DQTSCVRACPPDKMEVD-KNGLKMCEPCGGLCPKACEGTG-----SGSRFQTVDSSNIDG 350

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     :|||       | ||:| || |: |||||| | :||  : : |
Db        351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVFRTVREITGYLNIQSWPPHMHNFS 410

Qy        424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
              || ||  | || |:|  :| | :: | :: || |||:|: :|   |  |  ||: |:: |
Db        411 VFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNW 470

Qy        483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECV 541
              ::  | | :  |   |||  :|| ||  |  ||: | |||||| ||::|  : ||  ||
Db        471 TKVLRGPTEERLDIKHNRPRRDCVAEGKVCDPLCSSGGCWGPGPGQCLSCRNYSRGGVCV 530

Qy        542 EECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCP 601
                 | | |||: :   |  |||||||  |: || |  :| |  |||::| | ||: ||
Db        531 THCNFLNGEPREFAHEAECFSCHPECQPMEGTATCNGSGSDTCAQCAHFRDGPHCVSSCP 590

Qy        602 FNNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAEQRA----SPL 657
                                :  || :  |:||  |||  |   : : |  :       : |
Db        591 HGVLGAK--------GPIYKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLIGKTHL 642

Qy        658 TSIVSAVVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGAMPNQA 716
              |  :: : |  |||:  :: |  :  | ::|: |  ||| |:  | :|||  ||    |:
Db        643 TMALTVIAG--LVVIFMMLGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKV 699

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
                || |||||||:|||||| ||||:||:|||:||::|||| |||: : :  :: : : |
Db        700 LARIFKETELRKLKVLGSGVFGTVHKGVWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDH 759

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
                :  :   :: ||||:|  |::|||||  :| | ||||||:|| || | ||||| :|||
Db        760 MLAIGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVRQHRGALGPQLLLNWGVQIA 819

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              |||  |||:  :|||:|||||||:|||: |:: |||:| ||   |: :       | ||||||
Db        820 KGMYYLEEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKWMA 879

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||   ::|||||||||||||||||||:|| |:   |:||||||||| || |||||
Db        880 LESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEVPDLLEKGERLAQPQICTID 939

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPA---SPLDSTFYRS 1013
              |||:|||||||   || |:|| :||:|||||| |::||: |  ||     |
Db        940 VYMVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNK 998

Qy       1014 LLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP-S 1072
                ||: ::   :| : |  ::                        : :|: : |  |:| :   :
Db        999 KLEEVELEPELDLDLDLEAEED------------------NLATTTLGSALSLPVGTLN 1039

Qy       1073 EEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHD-PSPLQRYSEDPTVPLP---- 1127
                  :| |:|| |    :  |:||    :  |  :  | |:  :      |:|
Db       1040 RPRGSQSLLSPSSGY-MPMNQGNLGESCQESAVSGSSERCPRPVSLH------PMPRGCL 1092

Qy       1128 --SETDGYVA----------------PLTCSPQPE----YVNQPDVRPQPPSPREGP--- 1162
                  ::|:|                : ||:|    | :|      | :|    |
Db       1093 ASESSEGHVTGSEAELQEKVSMCRSRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPPGLE 1152

Qy       1163 --------LPAARPAGATLERAKTLSP-GKNGVV------KDVFAFGGAVENPEYLTPQG 1207
                       :|    |   |  |||  | : |:      :|         |  ||:  :
Db       1153 EEDVNGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEEEDED--------EEYEYMNRRR 1204

Qy       1208 GAAPQPHPPPAFSPAFDNLYYWD-----------QDPPERGAPPSTFKGTPTAENPEYL 1255
                :| ||||   |      | |           | |   |      ||   |: ||:
Db       1205 RHSP-PHPPRPSSLEELGYEYMDVGSDLSASLGSTQSCPLHPVPIMPTAGTTPDEDYEYM 1263


RESULT 10
ERB3_RAT
ID   ERB3_RAT       STANDARD;      PRT;  1339 AA.
AC   Q62799; Q62955;
```

```
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3).
GN   ERBB3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96096535; PubMed=8522190;
RA   Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT   "Cloning of the rat ErbB3 cDNA and characterization of the
RT   recombinant protein.";
RL   Gene 165:279-284(1995).
RN   [2]
RP   REVISIONS TO 85; 513 AND 565.
RA   Hellyer N.J., Koland J.G.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 922-1097 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U29339; AAC28498.2; -.
DR   EMBL; U52530; AAC53050.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1339       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN       20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     644    662       POTENTIAL.
FT   DOMAIN       663   1339       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       183    259       CYS-RICH.
FT   DOMAIN       707    964       PROTEIN KINASE.
FT   NP_BIND      713    721       ATP (BY SIMILARITY).
FT   BINDING      740    740       ATP (BY SIMILARITY).
FT   ACT_SITE     832    832       BY SIMILARITY.
FT   DISULFID     186    194       BY SIMILARITY.
FT   DISULFID     190    202       BY SIMILARITY.
FT   DISULFID     210    218       BY SIMILARITY.
FT   DISULFID     214    226       BY SIMILARITY.
FT   DISULFID     227    235       BY SIMILARITY.
FT   DISULFID     231    243       BY SIMILARITY.
FT   DISULFID     246    255       BY SIMILARITY.
FT   DISULFID     259    286       BY SIMILARITY.
FT   DISULFID     290    301       BY SIMILARITY.
FT   DISULFID     305    320       BY SIMILARITY.
FT   DISULFID     323    327       BY SIMILARITY.
FT   DISULFID     500    509       BY SIMILARITY.
FT   DISULFID     504    517       BY SIMILARITY.
FT   DISULFID     520    529       BY SIMILARITY.
FT   DISULFID     533    549       BY SIMILARITY.
FT   DISULFID     556    573       BY SIMILARITY.
FT   DISULFID     576    585       BY SIMILARITY.
FT   DISULFID     589    610       BY SIMILARITY.
FT   DISULFID     613    621       BY SIMILARITY.
FT   DISULFID     617    629       BY SIMILARITY.
FT   CARBOHYD     126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    1028   1028       L -> P (IN REF. 3).
SQ   SEQUENCE   1339 AA;  147545 MW;  0AA5F2402BBFDF1E CRC64;

  Query Match             33.9%;  Score 2316.5;  DB 1;  Length 1339;
  Best Local Similarity   40.4%;  Pred. No. 1.1e-116;
  Matches  521;  Conservative  168;  Mismatches  437;  Indels  165;  Gaps   35;

Qy          3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              |  ||     |  |  |:    |  ||  ||   | :    :     |  ||: |:|| ||
Db          7 LQVLC----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLYKLYEKCEVVMGN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:     || ||||| |:|| ||||:|  |:   :||  ||:||||||:::  :|: |:
Db         63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLPNLRVVRGTQVYDGKFAIFVM-- 120

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 ||  |      |   ||:|:   ||||| ||| |::| :||: ||| |:||
Db        121 ---LNYNT----NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRVR--- 170

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
                ::  |    | ||  :||| ||||   :||| ||:|:||  |  || || |  |||:
Db        171 GAEIVVKNNGANCPPCHEVCKG-RCWGPGPDDCQILTKTICAPQCNGRCFGPNPNQCCHD 229

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              :|| ||:||: :|| ||  || || |   ||  : ||  ||:  |||  :| :|  || :
Db        230 ECAGGCSGPQDTDCFACRRFNDSGACVPRCPEPLVYNKLTFQLEPNPHTKYQYGGVCVAS 289

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGL--GMQYIKANSKF 356
              ||:|::  |    |  ||    ||  : | : || |   | : | |   | :|   :|
Db        290 CPHNFV-VDQTFCVRACPPDKMEVD-KHGLKMCEPCGGLCPKACEGTGSGSRYQTVDSSN 347

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
              |       | | || |:| ||    : ||       | ||:| || |: |||||| | :||
Db        348 ID----GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLNVFRTVREITGYLNIQSWP 403

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLC 475
```

```
Db        404 PHMHNFSVFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLC 463

Qy        476 FVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQF 534

Db        464 YHHSLNWTRLLRGPSEERLDIKYDRPLGECLAEGKVCDPLCSSGGCWGPGPGQCLSCRNY 523

Qy        535 LRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPP 594

Db        524 SREGVCVTHCNFLQGEPREFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAHFRDGP 583

Qy        595 FCVARCPFNNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSC--VDLDDKGCPAEQ 652

Db        584 HCVNSCPHGILGAK--------GPIYKYPDAQNECRPCHENCTQGCNGPELQDCLGQAEV 635

Qy        653 RASPLTSIVSAVVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGA 711

Db        636 LMSKPHLVIAVTVG--LAVILMILGGSFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-E 692

Qy        712 MPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771

Db        693 KANKVLARIFKETELRKLKVLGSGVFGTVHKGIWIPEGESIKIPVCIKVIEDKSGRQSFQ 752

Qy        772 EILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNW 831

Db        753 AVTDHMLAVGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVKQHRETLGPQLLLNW 812

Qy        832 CMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVP 891

Db        813 GVQIAKGMYYLEEHSMVHRDLALRNVMLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTP 872

Qy        892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPP 951

Db        873 IKWMALESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPQ 932

Qy        952 ICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFY 1011

Db        933 ICTIDVYMVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAE 989

Qy       1012 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1071

Db        990 PSVLTTKEL-----QEAELEPEL----------------------------DLDLDLEA 1015

Qy       1072 SEE--------------------EAPRSPLAPSEG--------AGSDVFDGDLGMGAAKG 1103

Db       1016 EEEGLATSLGSALSLPTGTLTRPRGSQSLLSPSSGYMPMNQSSLGEACLDSAVLGGREQF 1075

Qy       1104 LQSLPTHDPSPLQRYSEDPTVPLPSETDGYV----APL-----TC-------SPQPE--- 1144

Db       1076 SRPISLH-PIPRGR-------PASESSEGHVTGSEAELQEKVSVCRSRSRSRSPRPRGDS 1127

Qy       1145 -YVNQPDVRPQPPSPREGP-----------LPAARPAGATLERAKTLSP-GKNGVV---- 1187

Db       1128 AYHSQRHSLLTPVTPLSPPGLEEEDGNGYVMPDTHLRGASSSREGTLSSVGLSSVLGTEE 1187

Qy       1188 --KDVFAFGGAVENPEYLTPQGGAAPQPHPP 1216

Db       1188 EDED--------EEYEYMNRKRRGSP-PRPP 1209


RESULT 11
EGFR_DROME
ID   EGFR_DROME     STANDARD;      PRT;  1426 AA.
AC   P04412; O61601; Q9W2G0; P81868;
DT   13-AUG-1987 (Rel. 05, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE   (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN   EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX   MEDLINE=94350209; PubMed=8070664;
RA   Clifford R., Schupbach T.;
RT   "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT   that several genetically defined classes of alleles cluster in
RT   subdomains of the receptor protein.";
RL   Genetics 137:531-550(1994).
RN   [2]
RP   REVISIONS.
RA   Clifford R., Schupbach T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=85124611; PubMed=2982499;
RA   Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT   "The Drosophila EGF receptor gene homolog: conservation of both
RT   hormone binding and kinase domains.";
RL   Cell 40:599-607(1985).
RN   [4]
RP   SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC   STRAIN=Oregon-R; TISSUE=Embryo;
RX   MEDLINE=87002474; PubMed=3093080;
RA   Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT   "Alternative 5' exons and tissue-specific expression of the
RT   Drosophila EGF receptor homolog transcripts.";
RL   Cell 46:1091-1101(1986).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=99102120; PubMed=9882502;
RA   Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT   "Several levels of EGF receptor signaling during photoreceptor
RT   specification in wild-type, Ellipse, and null mutant Drosophila.";
RL   Dev. Biol. 205:129-144(1999).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
```

```
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [7]
RP   SEQUENCE OF 959-1078 FROM N.A.
RC   STRAIN=Daekwanryeong;
RX   MEDLINE=85137938; PubMed=2983232;
RA   Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT   "A Drosophila genomic sequence with homology to human epidermal
RT   growth factor receptor.";
RL   Nature 314:178-180(1985).
RN   [8]
RP   SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=92038942; PubMed=1936959;
RA   Raz E., Schejter E.D., Shilo B.Z.;
RT   "Interallelic complementation among DER/flb alleles: implications for
RT   the mechanism of signal transduction by receptor-tyrosine kinases.";
RL   Genetics 129:191-201(1991).
RN   [9]
RP   REVIEW.
RX   MEDLINE=97248481; PubMed=9094709;
RA   Perrimon N., Perkins L.A.;
RT   "There must be 50 ways to rule the signal: the case of the Drosophila
RT   EGF receptor.";
RL   Cell 89:13-16(1997).
CC   -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC       WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC       MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC       CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC       DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC       POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC       ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC       AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC       SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC       CUTICLE.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC       MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC       PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC       TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC       UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC       TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC       EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC       IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC       POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC       AND THORACIC AND ABDOMINAL GANGLIA.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF052754; AAC08536.1; -.
DR   EMBL; AF052753; AAC08536.1; JOINED.
DR   EMBL; AF052754; AAC08535.1; -.
DR   EMBL; AF052752; AAC08535.1; JOINED.
DR   EMBL; K03054; AAA51462.1; -.
DR   EMBL; K03417; AAA51460.1; -.
DR   EMBL; K03416; AAA50965.1; -.
DR   EMBL; K03418; AAA51461.1; -.
DR   EMBL; AF109077; AAD26134.1; -.
DR   EMBL; AF109078; AAD26132.1; -.
DR   EMBL; AF109082; AAD26132.1; JOINED.
DR   EMBL; AF109078; AAD26133.1; -.
DR   EMBL; AF109084; AAD26133.1; JOINED.
DR   EMBL; AF109079; AAD26130.1; -.
DR   EMBL; AF109081; AAD26130.1; JOINED.
DR   EMBL; AF109079; AAD26131.1; -.
DR   EMBL; AF109083; AAD26131.1; JOINED.
DR   EMBL; AF109080; AAD26135.1; -.
DR   EMBL; AE003454; AAF46732.1; -.
DR   EMBL; X02293; CAA26157.1; -.
DR   EMBL; X78920; CAA55523.1; -.
DR   EMBL; X78918; CAA55521.1; -.
DR   EMBL; X78919; CAA55522.1; -.
DR   PIR; A00640; GQFFE.
DR   HSSP; P11362; 1FGK.
DR   FlyBase; FBgn0003731; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 7.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW   Developmental protein.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1426       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    868       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    869    889       POTENTIAL.
FT   DOMAIN      890   1426       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      938   1198       PROTEIN KINASE.
FT   NP_BIND     944    952       ATP (BY SIMILARITY).
FT   BINDING     971    971       ATP (BY SIMILARITY).
FT   ACT_SITE   1063   1063       BY SIMILARITY.
FT   MOD_RES     902    902       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

  Query Match           28.7%;  Score 1960;  DB 1;  Length 1426;
  Best Local Similarity   32.5%;  Pred. No. 1.4e-97;
  Matches  467;  Conservative  183;  Mismatches  431;  Indels  354;  Gaps   40;

Qy         24 QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT-NASLSFLQDIQEVQG 82
              ::| ||  :| :|:: | |    ||  |  |  | |||:||:||  |  |||| :|:|| |
Db        100 KICIGTKSRLSVPSNKEHHYRNLRDRYTNCTYVDGNLKLTWLPNENLDLSFLDNIREVTG 159

Qy         83 YVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGL 137
              |:||:|  |::|   :|:|:|| ||      |: ||| |                :  :
Db        160 YILISHVDVKKVVFPKLQIIRGRTLFSLSVEEEKYALFV---------------TYSKM 203

Qy        138 RELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSP 197
              |::  | ::| | |     |  ||:  || | :|                |  | |  |
Db        204 YTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPKCHE 263

Qy        198 MCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDCLAC 255
               |    ||||  ::||  ::  |:  ||  || || | :|||  || |||||   ||:||
Db        264 SCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDCIAC 322

Qy        256 LHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVC 315
               :|    : :  || :  ||  |:    ||||:| :||:||  ||  :|  | |:|   |
Db        323 KNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACVRSC 381

Qy        316 PLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EPAGCKK 369
              |    :   |    |  |: || : |                  |:| |       |  |
Db        382 PQDKMDKGGE-----CVPCNGPCPKTC-------------PGVTVLHAGNIDSFRNCTV 422

Qy        370 IFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDSLPD 421
```

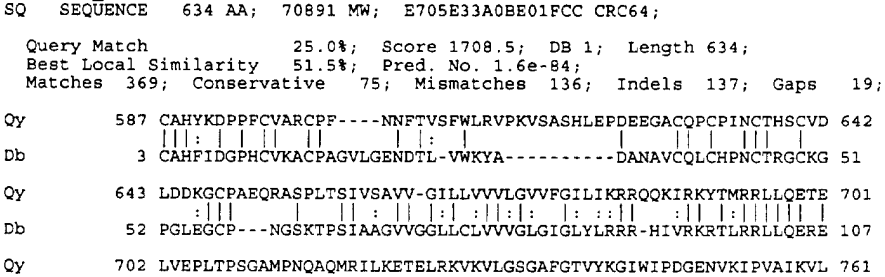


RESULT 12

ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC ---

CC ---
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.0%; Score 1708.5; DB 1; Length 634;
Best Local Similarity 51.5%; Pred. No. 1.6e-84;
Matches 369; Conservative 75; Mismatches 136; Indels 137; Gaps 19;



```
Db        108 LVEPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKEL 167

Qy        762 RENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRG 821

Db        168 REATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKD 227

Qy        822 RLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDET 881

Db        228 NIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEK 287

Qy        882 EYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLL 941

Db        288 EYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVL 347

Qy        942 EKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDL 1000

Db        348 EKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERM 407

Qy       1001 GPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRS 1060

Db        408 HLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-- 449

Qy       1061 GGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPL 1115

Db        450 ----------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFV 485

Qy       1116 QRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATL 1173

Db        486 QRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS---------------- 521

Qy       1174 ERAKTLSPGKNGVVKDVF---------------AFGGAVENPEYLTPQGGAAPQPHPPPA 1218

Db        522 -----TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPL 568

Qy       1219 FSPAFDNLYYWDQ--------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1260

Db        569 AKTVFESSPYWIQSGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 625
```

```
RESULT 13
ERBB_AVIER
ID   ERBB_AVIER     STANDARD;      PRT;   604 AA.
AC   P00535;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ES4).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=79685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H;
RX   MEDLINE=84026539; PubMed=6313229;
RA   Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT   "The erbB gene of avian erythroblastosis virus is a member of the src
RT   gene family.";
RL   Cell 35:71-78(1983).
RN   [2]
RP   SEQUENCE OF 1-152 FROM N.A.
RX   MEDLINE=84223957; PubMed=6328658;
RA   Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA   Saule S., Martin P., Stehelin D.;
RT   "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT   new type of oncogene.";
RL   Science 224:1456-1459(1984).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC       ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC       IN CHICKENS.
CC   -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC       RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02006; AAA42394.1; ALT_INIT.
DR   EMBL; K01216; AAA42400.1; -.
DR   PIR; A00644; TVYUH.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   CONFLICT     29     29       R -> W (IN REF. 2).
FT   CONFLICT    140    140       S -> F (IN REF. 2).
FT   CONFLICT    146    146       I -> V (IN REF. 2).
SQ   SEQUENCE   604 AA;  67633 MW;  76EBCDD06745D609 CRC64;

  Query Match             24.5%;  Score 1677;  DB 1;  Length 604;
  Best Local Similarity   51.6%;  Pred. No. 7.3e-83;
  Matches  361;  Conservative   73;  Mismatches  126;  Indels  140;  Gaps   18;

Qy        587 CAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVD 642

Db          3 CAHFIDGPHCVKACPAGVLGENDTL-----VRKYA------DANAVCQLCHPNCTRGCKG 51

Qy        643 LDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETE 701

Db         52 PGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERE 107

Qy        702 LVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVL 761

Db        108 LVEPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKEL 167

Qy        762 RENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRG 821

Db        168 REATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKD 227

Qy        822 RLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDET 881

Db        228 NIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEK 287

Qy        882 EYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLL 941

Db        288 EYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVL 347

Qy        942 EKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDL 1000

Db        348 EKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERM 407

Qy       1001 GPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRS 1060

Db        408 HLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-- 449

Qy       1061 GGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPL 1115
```

```
Db        450 ----------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFV 485

Qy       1116 QRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATL 1173

Db        486 QRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPSTAM------------- 524

Qy       1174 ERAKTLSPGKNGVVKDVFAF------------------GGAVENPEYLTPQGGAAPQPH 1214

Db        525 ------------VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTN 564

Qy       1215 PPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254

Db        565 QSPLAKTVFESSPYWIQSGNHQ----------INLDNPDY 594


RESULT 14
ERBB_AVIEU
ID   ERBB_AVIEU     STANDARD;      PRT;   540 AA.
AC   P11273;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ts167).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=103898;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87064458; PubMed=2878364;
RA   Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT   "A single amino acid substitution in v-erbB confers a thermolabile
RT   phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT   cells.";
RL   Mol. Cell. Biol. 6:1751-1759(1986).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13179; AAA42401.1; -.
DR   PIR; A25231; TVFVEB.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   VARIANT     270    270       H -> D (IN THERMOLABILE V-ERBB).
SQ   SEQUENCE   540 AA;  60412 MW;  5B53297AA068B65D CRC64;

  Query Match             23.2%;  Score 1589;  DB 1;  Length 540;
  Best Local Similarity   54.1%;  Pred. No. 3.3e-78;
  Matches  340;  Conservative   65;  Mismatches  120;  Indels  104;  Gaps   16;

Qy        587 CAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVD 642

Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYA----------DANAVCQLCHPNCTRGCKG 51

Qy        643 LDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETE 701

Db         52 PGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERE 107

Qy        702 LVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVL 761

Db        108 LVEPLTPSGEAPNQAHLRILKETEFKKVKVLGPGAFGTVYKGLWIPEGEKVTIPVAIKEL 167

Qy        762 RENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRG 821

Db        168 REATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKD 227

Qy        822 RLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDET 881

Db        228 NIGSQYLLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAKQLGADEK 287

Qy        882 EYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLL 941

Db        288 EYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVL 347

Qy        942 EKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDL 1000

Db        348 EKGERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERM 407

Qy       1001 GPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRS 1060

Db        408 HLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-- 449

Qy       1061 GGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPL 1115

Db        450 ----------------SRTPLLSSLSATSNNSATNCIDRNGG------------H----- 476

Qy       1116 QRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LE 1174

Db        477 -----------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAIS 517

Qy       1175 RAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203

Db        518 KLPIDSRYQN-------SHSTAVDNPEYL 539


RESULT 15
EGFR_CHICK
ID   EGFR_CHICK     STANDARD;      PRT;   703 AA.
AC   P13387;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE   (Fragment).
GN   EGFR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88261272; PubMed=3260329;
RA   Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA   Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT   "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT   expression in mouse cells, and differential binding of EGF and
RT   transforming growth factor alpha.";
RL   Mol. Cell. Biol. 8:1970-1978(1988).
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
```

```
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20386; AAA48760.1; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 4.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     30
FT   CHAIN        31   >703       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    667       POTENTIAL.
FT   DOMAIN      668   >703       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    197    206       BY SIMILARITY.
FT   DISULFID    201    214       BY SIMILARITY.
FT   DISULFID    222    230       BY SIMILARITY.
FT   DISULFID    226    238       BY SIMILARITY.
FT   DISULFID    239    247       BY SIMILARITY.
FT   DISULFID    243    255       BY SIMILARITY.
FT   DISULFID    258    267       BY SIMILARITY.
FT   DISULFID    271    298       BY SIMILARITY.
FT   DISULFID    302    314       BY SIMILARITY.
FT   DISULFID    318    333       BY SIMILARITY.
FT   DISULFID    336    340       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    581       BY SIMILARITY.
FT   DISULFID    569    589       BY SIMILARITY.
FT   DISULFID    592    601       BY SIMILARITY.
FT   DISULFID    605    627       BY SIMILARITY.
FT   DISULFID    630    638       BY SIMILARITY.
FT   DISULFID    634    646       BY SIMILARITY.
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    200    200       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    359    359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    420    420       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    633    633       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    648    648       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     703    703
SQ   SEQUENCE   703 AA;  77427 MW;  AFF2DE11B735A690 CRC64;

  Query Match             22.7%;  Score 1554;  DB 1;  Length 703;
  Best Local Similarity   43.9%;  Pred. No. 3.3e-76;
  Matches  315;  Conservative  107;  Mismatches  252;  Indels   44;  Gaps   16;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              :|   :| ||  | |        :|| ||: ||      | |   |: :|  |:|| |||

Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  |:||: |||| |||||| | |  :||: |:|:||  |:::::|||| |
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
              :| |          ||||| :: |:||| ||| |  ||:||  ||:|| ||   : :  |
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              |::|  :   :|  | | |    |||   ::||:||: :||  |: ||:| :|:|||| |
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||: ||||| |      |:  || || ||  |::   ||||:|:|||:|| |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||   |    ||  |:| ::|:||   |::|| |:|:  :|    |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  | || | :: || :| ||  :  | || |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
              ||  |:||::|||   :| ||| :  | |  ||||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: |  ||    |      || :::|  :   |  ||:   ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
              |||::| :||| |||:    ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAE 651
              ||  ||      | |:  |          |    || |  |||  |      :|||
Db        601 CVKACPAGVLGENDTL-VWKYA----------DANAVCQLCHPNCTRGCKGPGLEGCP-- 647

Qy        652 QRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTP 708
                |  || : || |:| :||:|:  |: ::||   :|| |:|||||| ||||||||
Db        648 -NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTP 703


Search completed: July 22, 2003, 08:47:01
Job time : 20.2793 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-579-593-14
Perfect score: 6841
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6114	89.4	1259	6	O18735	O18735 canis famil
2	3099	45.3	1209	11	Q9QX70	Q9qx70 rattus norv
3	3077	45.0	1210	11	Q9EP98	Q9ep98 mus musculu
4	2687	39.3	1165	13	Q9YH40	Q9yh40 xiphophorus
5	2667.5	39.0	1137	13	Q9W6F6	Q9w6f6 gallus gall
6	2259	33.0	1328	13	P79754	P79754 fugu rubrip
7	2026.5	29.6	1433	5	Q9BIH9	Q9bih9 anopheles g
8	1871	27.3	419	4	Q9UK79	Q9uk79 homo sapien
9	1739	25.4	367	11	Q8R2X1	Q8r2x1 mus musculu
10	1697.5	24.8	412	4	Q8WYV0	Q8wyv0 homo sapien
11	1679	24.5	729	15	Q86712	Q86712 avian rous-
12	1677	24.5	567	15	Q86714	Q86714 avian rous-
13	1612.5	23.6	962	15	Q64895	Q64895 avian eryth
14	1604	23.4	545	15	Q85468	Q85468 avian eryth
15	1468.5	21.5	655	11	Q9WVF5	Q9wvf5 mus musculu
16	1452.5	21.2	643	11	Q9ERV6	Q9erv6 mus musculu
17	1273	18.6	1193	5	Q9Y1X8	Q9y1x8 ephydatia f
18	1206.5	17.6	1368	5	Q23821	Q23821 caenorhabdi
19	1176	17.2	1717	5	Q26566	Q26566 schistosoma
20	1126	16.5	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.6	478	11	Q9ESE0	Q9ese0 rattus norv
22	906	13.2	165	4	Q14256	Q14256 homo sapien
23	901.5	13.2	599	13	Q9PSH2	Q9psh2 gallus gall
24	806.5	11.8	346	13	P11776	P11776 xiphophorus
25	783	11.4	176	11	Q923V5	Q923v5 rattus norv
26	778	11.4	435	5	Q8SZW1	Q8szw1 drosophila
27	754.5	11.0	311	13	Q99162	Q99162 xiphophorus
28	734	10.7	331	4	Q9BUD7	Q9bud7 homo sapien
29	731	10.7	1671	5	Q9NJV5	Q9njv5 biomphalari
30	723	10.6	149	6	Q9BG66	Q9bg66 oryctolagus
31	720.5	10.5	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
32	693	10.1	1418	13	O93457	O93457 scophthalmu
33	685.5	10.0	1368	13	Q8UW85	Q8uw85 paralichthy
34	670.5	9.8	1369	13	Q8UW86	Q8uw86 paralichthy
35	668.5	9.8	1472	5	Q9U5A8	Q9u5a8 bombyx mori
36	665	9.7	1358	13	O73798	O73798 xenopus lae
37	653	9.5	1412	13	Q8UW84	Q8uw84 paralichthy
38	644.5	9.4	1418	13	Q8UW83	Q8uw83 paralichthy
39	633	9.3	1245	13	Q9YGH8	Q9ygh8 scophthalmu
40	628	9.2	1371	11	Q9QVW4	Q9qvw4 rattus sp.
41	619.5	9.1	2144	5	Q9VD94	Q9vd94 drosophila
42	592	8.7	987	11	Q91YM0	Q91ym0 mus musculu
43	587.5	8.6	1036	4	Q07912	Q07912 homo sapien
44	587.5	8.6	1055	11	O54967	O54967 mus musculu
45	587	8.6	935	4	Q96L35	Q96l35 homo sapien

ALIGNMENTS

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErbB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB008451; BAA23127.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 89.4%; Score 6114; DB 6; Length 1259;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 1133; Conservative 44; Mismatches 66; Indels 34; Gaps 4;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||||||||||| ||| |||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGLLLALLPSGAAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| |||||||||||||||||||:||||:|||||||||||||||||||||||||||
Db         61 ELTYLPANASLSFLQDIQEVQGYVLIAHSQVRQIPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||      ||: |||||||||||||||||||||||:||||:|||||||||:||||||||
Db        121 DPLEGGIPAPGAAQGGLRELQLRSLTEILKGGVLIQRSPQLCHQDTILWKDVFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||  || |||| || : ||| || ||||||||||||||||||| |||||||||
Db        181 LTLIDTNRFSACPPCSPACKDAHCWGASSGDCQSLTRTVCAGGCARCKGPQPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||:|||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||:||| |||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLRVFEALEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              :||||||:|||||:||:|||||||||||||||||||||||||| |   |||||||
Db        421 NLSVFQNLRVIRGRVLHDGAYSLTLQGLGISWLGLRSLRELGSGLALIHRNARLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||:|||||:|||||||||:  || ||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHSANRPEEECVGEGLACYP-CAHGHCWGPGPTQCVNCSQFLRGQEC 539

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |||||||||||||||  |:||||| |||||||||||| ||||||||||||||||||||
Db        540 VEECRVLQGLPREYVKDRYCLPCHSECQPQNGSVTCFGSEADQCVACAHYKDPPFCVARC 599

Qy        601 PFNNFTVSFWLRVPKVSASHLEP-----------DEEGACQPCPINCTHSCVDLDDKGCP 649
              |                  | ::|          |||| |||||||||||| |||:||||
Db        600 P-----------------SGVKPDLSFMPIWKFADEEGTCQPCPINCTHSCADLDEKGCP 642

Qy        650 AEQRASPLTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPS 709
              |||||||:|||::|||||| ||:|:| |||||||:||||||||||||||||||||||||
Db        643 AEQRASPVTSIIAAVVGILLAVVVGLVLGILIKRRRQKIRKYTMRRLLQETELVEPLTPS 702

Qy        710 GAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 GAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKA 762

Qy        770 NKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLL 829
              ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        763 NKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLL 822

Qy        830 NWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGK 889
              |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 NWCVQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGK 882

Qy        890 VPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQ 949
              ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        883 VPIKWMALESIPPRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQ 942

Qy        950 PPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDST 1009
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        943 PPICTIDVYMIMVKCWMIDSECRPRFRELVAEFSRMARDPQRFVVIQNEDLGPASPLDST 1002

Qy       1010 FYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGL 1069
              |||||||||||||||||||||||||||||:| |||||  | ||||||||:|||:|||||
Db       1003 FYRSLLEDDDMGDLVDAEEYLVPQQGFFCPEPTPGAGGTAHRRHRSSSTRNGGGELTLGL 1062

Qy       1070 EPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSE 1129
              |||||| |:|||||||||||||||||||||||||||||: |||||||||||||||| |
Db       1063 EPSEEEPPKSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPSQDPSPLQRYSEDPTVPLPPE 1122

Qy       1130 TDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLER-----AKTLSPGKN 1184
              ||| |||||||||||||||:| ||||   ||||| :|||||||||      ||||||||
Db       1123 TDGKVAPLTCSPQPEYVNQPEVWPQPPLALEGPLPPSRPAGATLERPKTLSPKTLSPGKN 1182

Qy       1185 GVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFK 1244
              |||||||||| |||||||| |:| ||||||||||||||||||||||| |||:||||||:
Db       1183 GVVKDVFAFGSAVENPEYLAPRGRAAPQPHPPPAFSPAFDNLYYWDQDPSERGSPPSTFE 1242

Qy       1245 GTPTAENPEYLGLDVPV 1261
              |||||||||||||||||
Db       1243 GTPTAENPEYLGLDVPV 1259
```

RESULT 2
Q9QX70
ID Q9QX70 PRELIMINARY; PRT; 1209 AA.
AC Q9QX70;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=90258888; PubMed=2342466;
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue.";
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; M37394; AAF14008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CC1B7773 CRC64;

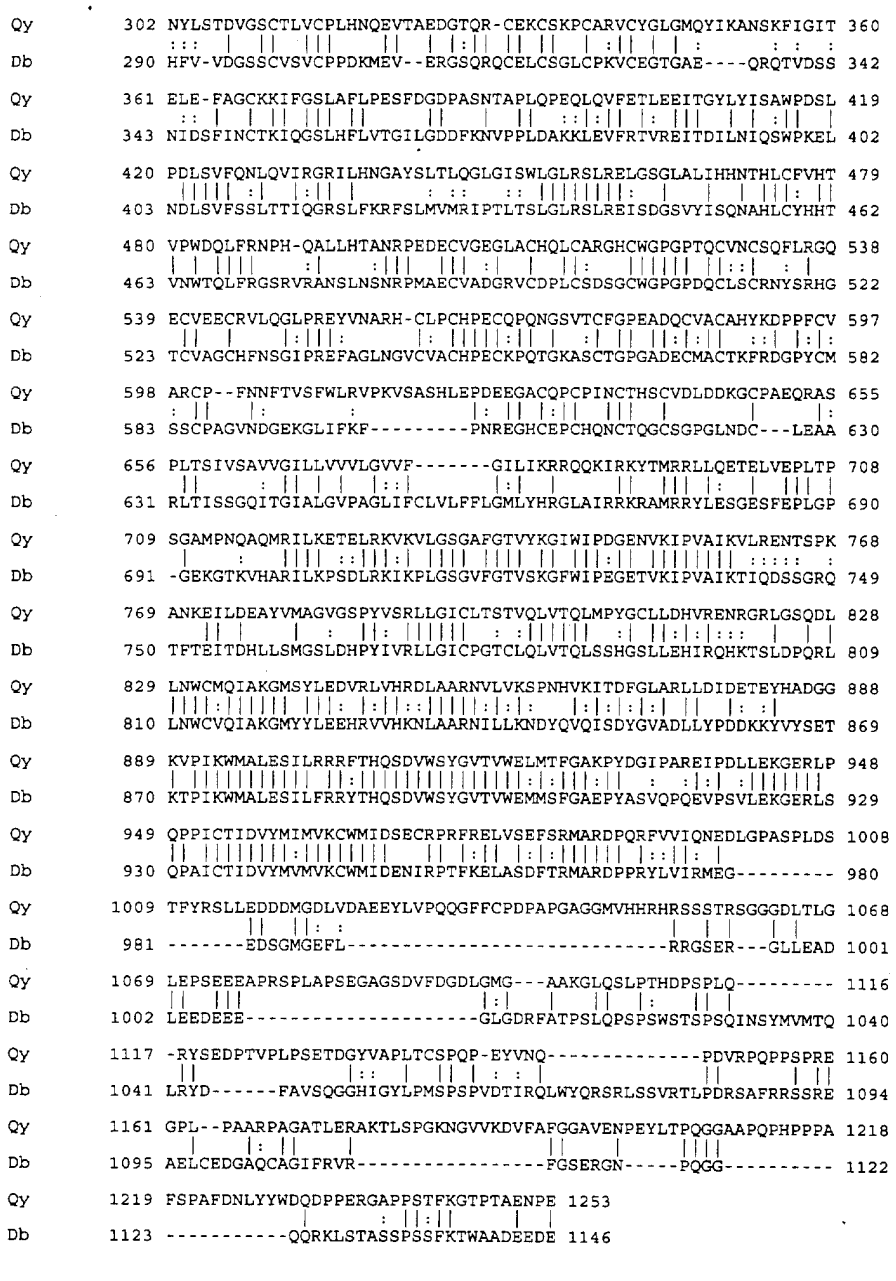
Query Match 45.3%; Score 3099; DB 11; Length 1209;
Best Local Similarity 49.9%; Pred. No. 1.7e-225;

```
  Matches  643;  Conservative  161;  Mismatches  353;  Indels  132;  Gaps   31;

Qy          3 LAALCRWGLLLALLPPGA-ASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE  61
              ||||| |          ||    :|| ||  :|    : | |    |: ::  |:|| ||||
Db         15 LAALCAAG--------GALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMFNNCEVVLGNLE  66

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  ||||: |||| |||||| | | ::||: |:|:||  |:|: |||||| |
Db         67 ITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAVLSN-- 124

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
                 | : |||||  :|:| ||| | |    || ||  :|| |:||  ::  |:
Db        125 --------YGTNKTGLRELPMRNLQEILIGAVRFSNNPILCNMETIQWRDIV-QDVFLSN 175

Qy        182 TLIDTNRS-RACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              :|  |    |  | | |    |||   |:|| ||: :||   |: ||:|  |:|||| |
Db        176 MSMDVQRHLTGCPKCDPSCPNGSCWGRGEENCQKLTKIICAQQCSRRCRGRSPSDCCHNQ 235

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:|| |
Db        236 CAAGCTGPRESDCLVCHRFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKC 295

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              | ||: || |||    |      ||  |||  :|:||  || :|| |:|:  | ::  |
Db        296 PRNYVVTDHGSCVRACGPDYYEV-EEDGVSKCKKCDGPCRKVCNGIGIGEFK-DTLSINA 353

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  |  | | |  || :| ||  : | || | :|:: :|::||||:| | |||::
Db        354 TNIKHFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPEN 413

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
                ||  |:||::|||   :| :|| : || |: ||||||:|:  |  :|  | :||: :
Db        414 WTDLHAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYAN 473

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: | :||  |:|      || | :|      |: ||:   |||| || ||:|    ||:
Db        474 TINWKKLFGTPNQKTKIMNNRAEKDCKATNHVCNPLCSSEGCWGPEPTDCVSCQNVSRGR 533

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||::| :|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||
Db        534 ECVDKCNILEGEPREFVENSECIQCHPECLPQTMNITCTGRGPDNCIKCAHYVDGPHCVK 593

Qy        599 RCPF-----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAEQR 653
                ||      || |  |    | :      |    |  | |||: |     ||| :|
Db        594 TCPSGIMGENNTLV--W----KFA------DANNVCHLCHANCTYGCAGPGLKGC--QQP 639

Qy        654 ASP-LTSIVSAVVGILLVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 711
              | : || : :|| || :|: |  || |  ||:| :|| |:|||| ||||||||||
Db        640 EGPKIPSIATGIVGGLLFIVV-VALGIGLFMRRRQLVRKRTLRRLLQERELVEPLTPSGE 698

Qy        712 MPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
              |||| :|||||||| :|:|||||||||||||||:|||:|| |||||||| ||| |||||||
Db        699 APNQAHLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANK 758

Qy        772 EILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNW 831
              |||||||||| | :|:| ||||||||||||:|||||||||||:|||::  :||| ||||
Db        759 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNW 818

Qy        832 CMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVP 891
              |:||||||:|||| ||||||||||||||:| ||||||||||:||  :| ||||:|||||
Db        819 CVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVP 878

Qy        892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPP 951
              |||||||||| | :|||||||||||||||||||:|||||||| ||  :||||||||||
Db        879 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPP 938

Qy        952 ICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
              |||||||||||||||:: ||:|||: |||:|||||||::||| :| :   || || |
Db        939 ICTIDVYMIMVKCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNF 998

Qy       1011 YRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLE 1070
              ||:|:|::|| |:|||:|||:|||||                | ||
Db        999 YRALMEEEDMEDVVDADEYLIPQQGFF---------------NSPST------------ 1030

Qy       1071 PSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET 1130
                    |:||  |  | |:       :         |    : : ||||| |||  |  :
Db       1031 ------SRTPLLSSLSANSN----SSTVACINRNGSCRVKEDAFLQRYSSDPTSVLTEDN 1080

Qy       1131 --DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVK 1188
                | ::      | |||:||  |  :|    : |:   :|             :||::  :
Db       1081 IDDTFL------PVPEYINQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQ 1125

Qy       1189 DVFAFGGAVENPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP----- 1233
              :      || ||||| | |         |   |  ||:  | |           :|
Db       1126 N--PHSNAVSNPEYLNTAQ---------PTCLSSGFDSSALWIQKGSHQMSLDNPDYQQD 1174

Qy       1234 --PERGAPPSTFKGTPTAENPEYLGLDVP 1260
                |:   |   ||| ||||| ||| :  |
Db       1175 FFPKEAKPNGIFKG-PTAENAEYLRVAPP 1202


RESULT 3
Q9EP98
ID   Q9EP98      PRELIMINARY;      PRT;  1210 AA.
AC   Q9EP98;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor isoform 1.
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative Egfr transcripts encoding truncated receptor
RT   isoforms.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative Egfr transcripts encoding truncated receptor
RT   isoforms.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF275366; AAG28045.1; -.
DR   EMBL; AF275364; AAG28045.1; JOINED.
DR   EMBL; AF275365; AAG28045.1; JOINED.
DR   EMBL; AF275367; AAG24386.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
```

```
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase.
SQ   SEQUENCE   1210 AA;  134840 MW;  62CD021C9DE32E18 CRC64;

  Query Match             45.0%;  Score 3077;  DB 11;  Length 1210;
  Best Local Similarity   49.5%;  Pred. No. 7.8e-224;
  Matches  635;  Conservative  160;  Mismatches  361;  Indels  128;  Gaps   27;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPF 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 -----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPL 657
Db        598 GIMGENNTLV--WKYA----------DANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKI 645

Qy        658 TSIVSAVVGILLVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
Db        646 PSIATGIVGGLLFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQA 704

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
Db        705 HLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDE 764

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
Db        765 AYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIA 824

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
Db        825 KGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMA 884

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
Db        885 LESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTID 944

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLL 1015
Db        945 VYMIMVKCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALM 1004

Qy       1016 EDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEE 1075
Db       1005 DEEDMEDVVDADEYLTPQQGFF----------------NSPST----------------- 1031

Qy       1076 APRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGY 1133
Db       1032 -SRTPLLSSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAF 1086

Qy       1134 VAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAF 1193
Db       1087 L------PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PH 1129

Qy       1194 GGAVENPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PER 1236
Db       1130 SNAVGNPEYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKE 1180

Qy       1237 GAPPSTFKGTPTAENPEYLGLDVP 1260
Db       1181 TKPNGIFKG-PTAENAEYLRVAPP 1203


RESULT 4
Q9YH40
ID   Q9YH40      PRELIMINARY;      PRT;  1165 AA.
AC   Q9YH40;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase proto-oncogene.
GN   XMRK.
OS   Xiphophorus xiphidium.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8086;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RX   MEDLINE=98241172; PubMed=9582016;
RA   Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA   Altschmied J., Schartl M.;
RT   "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT   overexpression and mutational alterations.";
RL   Oncogene 16:1681-1690(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53471; AAD10500.2; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
```

```
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1165 AA;  129614 MW;  7F7EE38D8771A74E CRC64;

  Query Match           39.3%;  Score 2687;  DB 13;  Length 1165;
  Best Local Similarity   45.1%;  Pred. No. 2.5e-194;
  Matches  579;  Conservative  161;  Mismatches  385;  Indels  160;  Gaps   32;

Qy          1 MELAALCRWGLLLALLPPG-AAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQV 55
              :||  |    ||| ||  |   ||     :|| ||  ::  :    : |   :: :| || |
Db          4 LELLEL----LLLLLLSIGRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNV 56

Qy         56 VQGNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALA 115
              |  |||:||   |  |||||  |||| |||||| |:|  :||  ||::||  :| |: |
Db         57 VLENLEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLL 116

Qy        116 VLDNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHK 175
              |: |    | ::|       ||::||| :|||| ||| :  || ||  :|| | || |
Db        117 VMSNYQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDK 173

Qy        176 NNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTD 234
              :   : ||      | |  | | |     ||       ||  |: :||  |  ||:|| | |
Db        174 TSNPTMNLIPHAFERQCQKCDPGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPID 233

Qy        235 CCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
              ||:| || |||||: :|||||  ||  | |:  ||    |:  : : : ||  :|||||:
Db        234 CCNEHCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAA 293

Qy        295 CVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGM----QYI 350
              ||  || ||: |: |:|   |     ||  |:| : |: |   | :|| |:|:      |
Db        294 CVKECPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSLSNTI 351

Qy        351 KANSKFIGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
              ||  ||     |: | || | :     ||:|||     |: || |    |::||||||
Db        352 AVNSTNIG----SFSNCTKINGDIILNRNSFEGDPHYKIGPMDPEHLWNLTTVKEITGYL 407

Qy        411 YISAWPDSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIH 469
              |  ||:::  |||||||::|||   :  :| : :|   : |||| ||:|: :|  ::
Db        408 VIMWWPENMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVSHLQWLGLSSLKEVSAGNVILK 467

Qy        470 HNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCV 529
              :   | :  |: | :|||:  |:: : |          |   |:  |:   |||||| ||
Db        468 NTPQLRYASTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCV 520

Qy        530 NCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAH 589
              :|   ||  || | :||| |||       |: || ||  |  :||:||     |  |||
Db        521 SCLHVDRGGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKCAH 580

Qy        590 YKDPPFCVARCP---FNNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDK 646
              ::| | |: |||     :     |           |: | ||||  |||  |
Db        581 FQDGPQCIPRCPHGMLGDGDTLIWKYA----------DKMGQCQPCHQNCTQGCSGPGLS 630

Qy        647 GCPAEQRASPLTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPL 706
              ||  :   |  :  |  | |:|: |:: ::  :|::||: | || |:|||||  ||||||
Db        631 GCRGD-IVSHSSLAVGLVSGLLITVIVALLIVVLLRRRRIK-RKRTIRRLLQEKELVEPL 688

Qy        707 TPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTS 766
              ||||  |||| :|||||| :| :|||||||||||||:| |||||::||||||||| ||
Db        689 TPSGQAPNQAFLRILKETEFKKDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATS 748

Qy        767 PKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQ 826
              || |:|:|||||||  |  |:| |||||||| |||||||||||||:||::: |:  |
Db        749 PKVNQEVLDEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQ 808

Qy        827 DLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHAD 886
              |||||:||||||:|||:  |||||||||||:|:|||||||||::||  || || |
Db        809 WLLNWCVQIAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQAH 868

Qy        887 GGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGER 946
              |||||||||||||:  :|||||||||||||||||:|||||||:||  :|| |||
Db        869 GGKVPIKWMALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGER 928

Qy        947 LPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPL 1006
              |||||||||:||||::||||||   |||||||| |||:|||| |::|||       ||
Db        929 LPQPPICTIEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSPS 985

Qy       1007 DSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLT 1066
              |   :  ||  ||  |:|||:|||:                          |::  : | |
Db        986 DRRLFSRLLSSDD--DVVDADEYLL--------------------RYKRIN-RQGS---- 1018

Qy       1067 LGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-P 1125
                          |  | |                          |  : |   ||  |||
Db       1019 ------------EPCIPPNGH--------------------PVRENSIALRYISDPTQNA 1046

Qy       1126 LPSETDGYVAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGAT 1172
              |  : ||:          ||||||              |:  :     ||:  ::: |
Db       1047 LEKDLDGH----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETN 1096

Qy       1173 LERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQD 1232
                | : |:  :| :         |::::|:|   | |          :  ||
Db       1097 FSRPEYLNTNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF--------- 1131

Qy       1233 PPERGAPPSTFKGTPTAENPEYLGL 1257
              |: ||        | ||| |||||
Db       1132 LPQTGALTGNGMFLPAAENLEYLGL 1156


RESULT 5
Q9W6F6
ID   Q9W6F6      PRELIMINARY;      PRT;  1137 AA.
AC   Q9W6F6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase (Fragment).
GN   ERBB4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HINDBRAIN;
RX   MEDLINE=99263203; PubMed=10328884;
RA   Dixon M., Lumsden A.;
RT   "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
RT   embryonic chick hindbrain.";
RL   Mol. Cell. Neurosci. 13:237-258(1999).
DR   EMBL; AF121963; AAD31764.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001368; TNFR_c6.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 1.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW   Kinase; Tyrosine-protein kinase.
FT   NON_TER       1       1
SQ   SEQUENCE   1137 AA;  127927 MW;  4D616436F87DC84F CRC64;

  Query Match           39.0%;  Score 2667.5;  DB 13;  Length 1137;
```

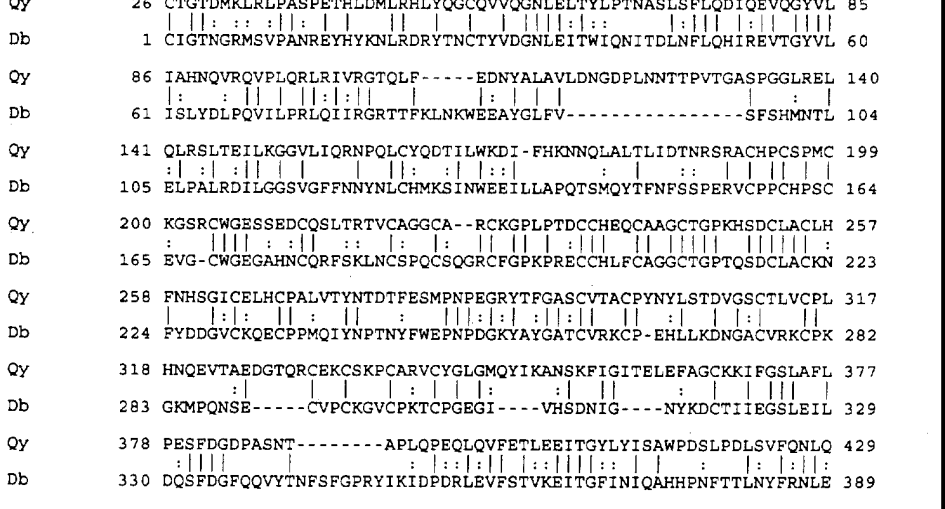
```
  Best Local Similarity   46.2%;  Pred. No. 7.2e-193;
  Matches  532;  Conservative  171;  Mismatches  353;  Indels   95;  Gaps   27;

Qy        161 LCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVC 220
              ||: ||| |:||          ||: || |   |   |    | | |||| :   ||:||:|||
Db          3 LCFADTIHWQDIVRNPWASNFTLVPTNGSSGCGRCHKSCTG-RCWGPTENHCQTLTKTVC  61

Qy        221 AGGC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTF 279
              |  |  || ||  :|||| :|| ||:||| :|| ||::|| || |   ||     ||  ||
Db         62 AEQCDGRCYGPYVSDCCHRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTF 121

Qy        280 ESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCA 339
              :   |   :||:|| ||  ||:|:::  |  ||   ||    ||  |:| : |: |:  |
Db        122 QLEHNHNAKYTYGAFCVKKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICP 179

Qy        340 RVCYGLGMQYIKANSKFIGITELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQ 398
              : | |:|   :  ::: :  : :: |  | || |:| ||     |||    | : ||:|
Db        180 KACDGIGTGSL-VSAQTVDSSNIDKFINCTKINGNLIFLVTGIHGDPYHTIAAINPEKLN 238

Qy        399 VFETLEEITGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSL 458
              :|:|: |||||| | :||::: |  || ||  | || |::|   | |:  ||: |  :||
Db        239 IFQTVREITGYLNIQSWPENMTDFRVFSNLVTIGGRALYSGLSLLILKQQGITSLQFQSL 298

Qy        459 RELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARG 518
              ::: :|   |  |::||: ||| |  ||  | |  :   |:  : |  :|: |::||:
Db        299 KQISAGNIYITDNSNLCYYHTVNWTSLFSTPSQKTVIHRNKKAENCTADGMVCNELCSSD 358

Qy        519 HCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCF 577
              |||||| ||::| :|:||: |:|  | :  |  ||: |   |: | |:|:  ::  :||:
Db        359 GCWGPGPDQCLSCKRFIRGRTCIESCNLYDGEFREFANGSVCMECDPQCEKMEDNMITCY 418

Qy        578 GPEADQCVACAHYKDPPFCVARCP-----FNNFTVSFWLRVPKVSASHLEPDEEGACQPC 632
              ||  | |  | |:|| | || :||      |:|   :              ||:  | ||
Db        419 GPGPDHCTKCFHFKDGPNCVEKCPDGLQGANSFIFKY-------------ADEDRECHPC 465

Qy        633 PINCTHSCVDLDDKGC-----------PAEQRASPLTSIVSAVV-GILLVVVLGVVFGIL 680
                 |||  |         |         |   | :||  | : |: |: ::|::|: | :
Db        466 HPNCTQGCRGPASHDCIYYPWTRQSTLPQHAR-TPL--IAAGVIGGLFIIVIMGLTFAVY 522

Qy        681 IKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTV 740
              ::|:  | :|  :||  | |||||||||||  |||||:||||||::|||||||||||
Db        523 VRRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTV 580

Qy        741 YKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTV 800
              |||||:|:|| ||||||||:|  | | |||| | :||| :||  :   |::  ||||:||: |:
Db        581 YKGIWVPEGETVKIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTI 640

Qy        801 QLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVK 860
              |||||||:|||||:|  |::  :||| |||||:|||||| ||| ||||||||||||||||
Db        641 QLVTQLMPHGCLLDYVHEHKDNIGSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVK 700

Qy        861 SPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWE 920
              |||||||||||||||: || ||:|||||:|||||||| |  |:|||||||||||||:||
Db        701 SPNHVKITDFGLARLLEGDEKEYNADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWE 760

Qy        921 LMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVS 980
              ||||| ||||||| ||||||||||||||||||||||||||||:||||||:: ||:|:|| :
Db        761 LMTFGGKPYDGIPTREIPDLLEKGERLPQPPICTIDVYMVMVKCWMIDADSRPKFKELAA 820

Qy        981 EFSRMARDPQRFVVIQNED-LGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCP 1039
              |||||||||||::||| :| :   || || |:::||:::|: |::|||||||| | |  |
Db        821 EFSRMARDPQRYLVIQGDDRMKLPSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIP 879

Qy       1040 DPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS--PLAP-SEGAGSDVFDGDL 1096
              |  :  :          | ||      |: |    ||    | || ::|| :::|:
Db        880 PPIYTSRTRIDSNRNQFVYRDGGYAAEQGV-PMPYRAPGCIIPEAPVAQGATAEIFEDTC 938

Qy       1097 GMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQP 1149
              |  :   :    : |  |||| |||| :|        : |||: |:   |: :|:|
Db        939 CNGTLRKQVATLAKEDSSTQRYSADPTVFIPERVIRGELDEDGYMTPMRDKPKTDYLNPV 998

Qy       1150 DVRPQPPSPREGPLPAA-RPAGATLERAKTLSPGKNGVVKDVF--------AFGGAVENP 1200
             .:  |     : | | |   |              :|     :| :         |   :||

Db        999 EENPFVSRRKNGDLQAVDNPEYHN-------APNGQPKAEDEYVNEPLYLNTFANTLENA 1051

Qy       1201 EYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGA--PPSTFKGTPT---------- 1248
              |||         : : |     ||||  ||:   | |     |   :   |
Db       1052 EYL--------KNNLPEKAKKAFDNPDYWNHSLPPRSTLQHPDYLQEYSTKYFYKQNGRI 1103

Qy       1249 ----AENPEYL 1255
                  |||||||
Db       1104 RPIVAENPEYL 1114


RESULT 6
P79754
ID   P79754      PRELIMINARY;      PRT;  1328 AA.
AC   P79754;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ErbB3.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99177347; PubMed=10077531;
RA   Gellner K., Brenner S.;
RT   "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT   rubripes.";
RL   Genome Res. 9:251-258(1999).
DR   EMBL; AF056116; AAC34391.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   1328 AA;  148613 MW;  A333039258B647E9 CRC64;

  Query Match             33.0%;  Score 2259;  DB 13;  Length 1328;
  Best Local Similarity   39.7%;  Pred. No. 7.7e-162;
  Matches  514;  Conservative  157;  Mismatches  422;  Indels  202;  Gaps   33;

Qy          9 WGLLLALLPP--GAASTQ----VCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLEL 62
              | |:|  :      |||:|    || ||   |    | |   :: :  |:||::: ||||:
Db          4 WRLILMCVASRLRAASSQTQEAVCPGTQNGLSSTGSQENQYNLNKDRYKGCEIIMGNLEI 63

Qy         63 TYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDP 122
              | : :|   |||: |:|| |||||| |  :::|| :||::||  |:|  :||:|  |
Db         64 TQIESNWDFSFLKTIREVTGYVLIAMNHFQEIPLGQLRVIRGNSLYERRFALSVFLN--- 120

Qy        123 LNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALT 182
                  |  |  | || :| | :||||| ||| |  |  | |   : |:||  :||   :
Db        121 ----YPKDG--PSGLNQLGLMNLTEILDGGVQIINNKYLRYGPWVYWRDII-RNNDAPIE 173

Qy        183 LIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCA 241
              :          ||      | |: |||    : || ||:|||| |  || |  | |||| :||
Db        174 IQFNGERGVCH---KSC-GNYCWGPGKDQCQILTKTVCAPQCNDRCFGTSPRDCCHIECA 229

Qy        242 AGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY 301
              ||| ||  :|| ||  || || |   ||  : ||  ||:   ||  :| :|: ||: ||
Db        230 AGCKGPLDTDCFACRLFNDSGACVPQCPQTLIYNKQTFQMETNPNAKYQYGSICVSQCPT 289
```



RESULT 7
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.6%; Score 2026.5; DB 5; Length 1433;
Best Local Similarity 32.8%; Pred. No. 3.5e-144;
Matches 474; Conservative 199; Mismatches 395; Indels 379; Gaps 39;



```
Qy        430 VIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLFRN 488
Db        390 VVGGRQLKENLFASVYIVKTSLKSLELKSLKRVNSGSIVILENSDLCFVEDIDWSEIKKS 449

Qy        489 PHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRVLQ 548
Db        450 SDHEVMVQKNRNATECHEEGMECSEQCSKAGCWGKGPEQCLECKNVKYKGKCLDSCK--- 506

Qy        549 GLPREY-VNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCP------ 601
Db        507 SLPRLYSVDSKTCGDCHQECKD-----FCYGPNEDNCGSCMNVKDGRFCVAECPTTKHAM 561

Qy        602 ----FNNFTVSFWLRVPK----------------VSASHLE---------PD-------- 624
Db        562 NGTCINCHKTCVGCRGPRDTIAPDGCISCDKAIIGSDAKIERCLMKDESCPDGYYSDYVL 621

Qy        625 -EEG------------------------------------------------------- 627
Db        622 QEEGPLKQLSGKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKKGEQCEDECPQDFYANEE 681

Qy        628 --ACQPCPINCT---------HSCVDL---------DD-------KGCPAEQ--------- 652
Db        682 TRICLPCHQECRGCHGLGDDHHECRNLKLFEGDPYDNATTFTCVSNCPASHPYKRFPQEA 741

Qy        653 ----------------RASPLTSIVSAVVGILLVVVLGVVFGI---LIKRRQQKIRKYTM 693
Db        742 GKIGPYCSADSMQSGLRIEPQTQVKIVMGSVMALILLCVVFGIAFVLFSRHKNKKDAVKM 801

Qy        694 RRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVK 753
Db        802 TMALAGCEDSEPLRPSNVGPNLTKLRIIKEAEIRRGGVLGMGAFGRVFKGVWMPEGESVK 861

Qy        754 IPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLL 813
Db        862 IPVAIKVLMEMSGSESSKEFLEEAYIMASVEHPNLLKLLAVCMTSQMMLITQLMPLGCLL 921

Qy        814 DHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLA 873
Db        922 DYVRNNKDKIGSKALLNWSTQIARGMAYLEERRLVHRDLAARNVLVQTPSCVKITVFGLA 981

Qy        874 RLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIP 933
Db        982 KLLDFDSDEYRAAGGKMPIKWLALECIRHRVFTSKSDVWAFGITIWELLTYGARPYENVP 1041

Qy        934 AREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFV 993
Db       1042 AKDVPELIEIGHKLPQPDICSLDVYCILLSCWVLDADARPTFKQLAETFAEKARDPGRYL 1101

Qy        994 VIQNEDLGPASPLDSTFYRSLLEDDDMGDLV---------------------------- 1024
Db       1102 MI---------PGDKFMRLPSYTNQDEKDLIRTLAPVAMAAAAAAAAAGASNVDVPSTIA 1152

Qy       1025 DAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS----- 1079
Db       1153 ETDEYLQPKTRPSIMLPGPSA----------------------VEPS-DEMPKSLRYCK 1188

Qy       1080 -PLAP---SEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVA 1135
Db       1189 DPLKPDDETDGHGKEV-----GVGGIR---------------------LNLPLDEDDYLM 1222

Qy       1136 PLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGG 1195
Db       1223 P-TCQSQ----NQS----------------------------TPG----YMDLIGVPA 1243

Qy       1196 AVENPEYL---------TPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGT 1246
Db       1244 SVDNPEYLMGSTQAIAGLAQGSMG--PHTPP--------------------PPNTPNGM 1280

Qy       1247 PTAENPE 1253
Db       1281 PTHQHSQ 1287
```

RESULT 8
Q9UK79
ID   Q9UK79   PRELIMINARY;   PRT;   419 AA.
AC   Q9UK79;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Herstatin.
GN   HER-2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99415951; PubMed=10485918;
RA   Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT   "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT   autoinhibitor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF177761; AAD56009.2; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR002174; Furin-like.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 1.
DR   SMART; SM00261; FU; 1.
SQ   SEQUENCE   419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

  Query Match   27.3%; Score 1871; DB 4; Length 419;
  Best Local Similarity   98.8%; Pred. No. 3.3e-133;
  Matches 341; Conservative   1; Mismatches   3; Indels   0; Gaps   0;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGL 345
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARGTHSL 345
```

RESULT 9
Q8R2X1
ID   Q8R2X1   PRELIMINARY;   PRT;   367 AA.
AC   Q8R2X1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 40.2 kDa protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027080; AAH27080.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   367 AA;  40163 MW;  0BE03395F9E101B0 CRC64;

  Query Match             25.4%;  Score 1739;  DB 11;  Length 367;
  Best Local Similarity   88.0%;  Pred. No. 2.6e-123;
  Matches  323;  Conservative   11;  Mismatches   33;  Indels    0;  Gaps    0;

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||:||:||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSL 120

Qy       1015 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1074
              |||||||:|||||||||||||||| |||| | |   | |||||| ||||||:|||||||||
Db        121 LEDDDMGELVDAEEYLVPQQGFFSPDPALGTGSTAHRRHRSSSARSGGGELTLGLEPSEE 180

Qy       1075 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1134
              | |||||||||||||||||||| :| |||||  || ||||||||||:||| ||||||
Db        181 EPPRSPLAPSEGAGSDVFDGDLAVGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYV 240

Qy       1135 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1194
              ||| ||||||||||:|||| |   ||| |  |||||||| ||||||||||||||||||||
Db        241 APLACSPQPEYVNQPEVRPQSPLTPEGPPPPIRPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1195 GAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254
              |||||||| |: | | |||| ||||||||||||||:  |:| |||||:||||||||||
Db        301 GAVENPEYLAPRAGTASQPHPSPAFSPAFDNLYYWDQNSSEQGPPPSTFEGTPTAENPEY 360

Qy       1255 LGLDVPV 1261
              |||||||
Db        361 LGLDVPV 367


RESULT 10
Q8WYV0
ID   Q8WYV0      PRELIMINARY;      PRT;   412 AA.
AC   Q8WYV0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 44.7 kDa protein.
GN   PP3659.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA   Wan D.F., Gu J.R.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF318349; AAL55856.1; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein.

SQ   SEQUENCE   412 AA;  44702 MW;  034397FF3F27D2BC CRC64;

  Query Match             24.8%;  Score 1697.5;  DB 4;  Length 412;
  Best Local Similarity   80.5%;  Pred. No. 4.4e-120;
  Matches  330;  Conservative    5;  Mismatches   22;  Indels   53;  Gaps    4;

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 120

Qy       1015 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 180

Qy       1075 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 240

Qy       1135 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1194
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        241 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1195 GAVENPEYLTPQGGAAPQP----------------------------HPPPA---FSPAFDNL 1226
              |||||||||||||||| |                            || |:        :
Db        301 GAVENPEYLTPQGGAALSPTLLLPSAQPSTTSITGTRTHQSGGLHPAPSKGHLRQRTQST 360

Qy       1227 YYWD-QDPPER-----------------------GAPPSTFKGTPTAEN 1251
              : |  |  ||                        ||:| :||  | |
Db        361 WVWTCQCEPEGQVRRSPDVSSGSREGLTSAGIKRWEGPPTTSRGTCHARN 410


RESULT 11
Q86712
ID   Q86712      PRELIMINARY;      PRT;   729 AA.
AC   Q86712;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyprotein.
GN   POLYPROTEIN.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60725.1; -.
DR   HSSP; P03322; 1A6S.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR004028; Retro_M.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02813; Retro_M; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   729 AA;  80649 MW;  84D2F6914EFE1D63 CRC64;

  Query Match             24.5%;  Score 1679;  DB 15;  Length 729;
  Best Local Similarity   53.8%;  Pred. No. 2.6e-118;
```

```
  Matches  357;  Conservative   71;  Mismatches  119;  Indels  116;  Gaps   17;

Qy        569 PQNGSVTCFGPEADQCVACAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPD 624
              |:  :    ||  | |: |||: | | ||  ||      | |:  |           |
Db        141 PEETATPKTGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYA----------D 187

Qy        625 EEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKR 683
                || |  ||| |     :|||     |   || : || |:| :||:|:  |: ::|
Db        188 ANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRR 244

Qy        684 RQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKG 743
              |   :|| |:|||||| |||||||||  |||| :|||||| :|||||||||||||||
Db        245 R-HIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKG 303

Qy        744 IWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLV 803
              :|||:|| |||||||| ||| |||||||||||||||| | :| | ||||||||||||:
Db        304 LWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLI 363

Qy        804 TQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPN 863
              ||||||||||::||::  :||| |||||:||||||:|||: |||||||||||||||:|
Db        364 TQLMPYGCLLDYIREHKDNIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQ 423

Qy        864 HVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMT 923
              ||||||||||:||  || ||||:|||||||||||||| | :|||||||||||||||||||
Db        424 HVKITDFGLAKLLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMT 483

Qy        924 FGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFS 983
              ||:|||||||| ||  :|||||||||||||||||||||||||:: ||:||||::|||
Db        484 FGSKPYDGIPASEISSVLEKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFS 543

Qy        984 RMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPA 1042
              :||||| |::||| :| :   || || |||:|:|::|| |:|||:||||| ||||
Db        544 KMARDPPRYLVIQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----- 598

Qy       1043 PGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLG 1097
                  | ||                |:||   | |  : ::  |
Db        599 -----------NSPST-----------------SRTPLLSSLSATSNNSATNCID---- 625

Qy       1098 MGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQP 1155
                  :  |  |  : | :|||| |||     |:  ||::      | ||||||  : |:
Db        626 ----RNGQGHPVREDSFVQRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKK 673

Qy       1156 PSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVF---------------AFGGAVENP 1200
              ||                    :  :| :  ::                :   ||:||
Db        674 PS--------------------TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNP 712

Qy       1201 EYL 1203
              |||
Db        713 EYL 715


RESULT 12
Q86714
ID   Q86714      PRELIMINARY;      PRT;   567 AA.
AC   Q86714;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   V-erbB protein (Fragment).
GN   V-ERBB.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60727.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Tyrosine-protein kinase.
FT   NON_TER       1       1
SQ   SEQUENCE   567 AA;  63390 MW;  C6D9CBA7ADF725E1 CRC64;

  Query Match             24.5%;  Score 1677;  DB 15;  Length 567;
  Best Local Similarity   54.4%;  Pred. No. 2.5e-118;
  Matches  356;  Conservative   69;  Mismatches  113;  Indels  116;  Gaps   17;

Qy        578 GPEADQCVACAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCP 633
              ||  | |: |||: | | ||  ||      | |:  |           |   || |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYA----------DANAVCQLCH 47

Qy        634 INCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYT 692
                ||| |     :|||     |   || : || |:| :||:|:  |: ::||   :|| |
Db         48 PNCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRT 103

Qy        693 MRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENV 752
              :|||||| |||||||||  |||| :|||||| :|||||||||||||||:|||:|| |
Db        104 LRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKV 163

Qy        753 KIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCL 812
              ||||||| ||| |||||||||||||||| | :| | ||||||||||||:|||||||||
Db        164 KIPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCL 223

Qy        813 LDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGL 872
              ||::||::  :||| |||||:||||||:|||: |||||||||||||||:| |||||||||
Db        224 LDYIREHKDNIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGL 283

Qy        873 ARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 932
              |:||  || |||||:||||||||||||||| | :|||||||||||||||||:||||||
Db        284 AKLLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGI 343

Qy        933 PAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRF 992
              || ||  :|||||||||||||||||||||||||:: ||:||||::|||:|||||  |:
Db        344 PASEISSVLEKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRY 403

Qy        993 VVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHH 1051
              :||| :| :   || || |||:|:|::|| |:|||:||||| ||||
Db        404 LVIQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF-------------- 449

Qy       1052 RHRSSSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQS 1106
                | ||                |:||     | |  : ::  |        :  |
Db        450 --NSPST-----------------SRTPLLSSLSATSNNSATNCID--------RNGQG 481

Qy       1107 LPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLP 1164
                |  : | :|||| |||     |:  ||::       | ||||||  : |: ||
Db        482 HPVREDSFVQRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS------- 526

Qy       1165 AARPAGATLERAKTLSPGKNGVVKDVF---------------AFGGAVENPEYL 1203
                                :  :| :  ::                :   ||:|||||
Db        527 --------------TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 566


RESULT 13
Q64895
ID   Q64895      PRELIMINARY;      PRT;   962 AA.
AC   Q64895;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Gag,v-erb-A,v-erb-B protein.
GN   GAG,V-ERB-A,V-ERB-B.
OS   Avian erythroblastosis virus.
```

```
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90206603; PubMed=1969616;
RA   Bruskin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT   "Six amino acids from the retroviral gene gag greatly enhance the
RT   transforming potential of the oncogene v-erb-B.";
RL   Oncogene 5:15-24(1990).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; X52209; CAA36459.1; -.
DR   EMBL; X52211; CAA36459.1; JOINED.
DR   HSSP; P10828; 2NLL.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW   Transcription regulation; Transferase; Tyrosine-protein kinase;
KW   Zinc-finger.
SQ   SEQUENCE   962 AA;  108320 MW;  3C5AED791E4E95CE CRC64;

  Query Match             23.6%;  Score 1612.5;  DB 15;  Length 962;
  Best Local Similarity   50.9%;  Pred. No. 4.2e-113;
  Matches  357;  Conservative   69;  Mismatches  143;  Indels  133;  Gaps   20;

Qy        541 VEECRVLQGLPRE-YVNAR-HCLP------------------------------CHPECQ 568
              :|:|:    |  | |:| | | :|                              |  |
Db        354 IEKCQESYLLAFEHYINYRKHNIPHFWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413

Qy        569 PQNGSVTCFGPEADQCVACAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPD 624
              ||       ||  | |: ||| | | ||  ||      | |:  |                |
Db        414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYA----------D 455

Qy        625 EEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKR 683
                 || |  ||| |     :|||      |   || : || |:| :||:|:  |: ::|
Db        456 ANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRR 512

Qy        684 RQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKG 743
              |   :|| |:|||||| ||||||||||  |||| :|||||| :|||||| ||||||||
Db        513 R-HIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKG 571

Qy        744 IWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLV 803
              :|||:|| | |||||| ||| |||||||||||||||| | :|:| ||||||||||||:
Db        572 LWIPEGEKVTIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLI 631

Qy        804 TQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPN 863
              ||||||||||::||::  :||| ||||||:||||||:|||:  :|||||||||||||:|
Db        632 TQLMPYGCLLDYIREHKDNIGSQYLLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQ 691

Qy        864 HVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMT 923
              |||||||||: |  || |||||||||||||||||| | :|||||||||||||||||||
Db        692 HVKITDFGLAKQLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMT 751

Qy        924 FGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFS 983
              ||:|||||||| ||  :|||||||||||||||||||||||||  :: ||:||||::|||
Db        752 FGSKPYDGIPASEISSVLEKGERLPQPPICTIDVYMIMVKCWMSGADSRPKFRELIAEFS 811

Qy        984 RMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPA 1042
              :||||| |::||| :| :   || || |||:|:|::|| |:|||:||||| ||||
Db        812 KMARDPPRYLVIQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----- 866

Qy       1043 PGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAK 1102
                          | ||                    |:||  |  | |:          | |
Db        867 -----------NSPST-----------------SRTPLLSSLSATSN-------NSATK 890

Qy       1103 GLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGP 1162
              :   |                  |   ||::      | ||||||  : |: ||
Db        891 CIDRNGGH---------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQ 926

Qy       1163 LPAARPAGAT-LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203
                    | : :   | :|      :   ||:|||||
Db        927 NQIYNYISLTAISKLPMDSRYQN-------SHSTAVDNPEYL 961


RESULT 14
Q85468
ID   Q85468      PRELIMINARY;      PRT;   545 AA.
AC   Q85468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS   Avian erythroblastosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88217326; PubMed=2897102;
RA   Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT   "Common site of mutation in the erbB gene of avian erythroblastosis
RT   virus mutants that are temperature sensitive for transformation.";
RL   Oncogene Res. 1:265-278(1987).
DR   EMBL; X06943; CAA30024.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   545 AA;  60899 MW;  140DCE8CCA0F8AF4 CRC64;

  Query Match             23.4%;  Score 1604;  DB 15;  Length 545;
  Best Local Similarity   53.9%;  Pred. No. 8e-113;
  Matches  344;  Conservative   66;  Mismatches  122;  Indels  106;  Gaps   17;

Qy        578 GPEADQCVACAHYKDPPFCVARCPF----NNFTVSFWLRVPKVSASHLEPDEEGACQPCP 633
              ||  | |: |||: | | ||  ||      | |:  |                |  || |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYA----------DANAVCQLCH 47

Qy        634 INCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYT 692
                |||  |      :|||      |   || : || |:| :||:|:  |: ::||   :|| |
Db         48 PNCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRT 103

Qy        693 MRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENV 752
              :|||||| |||||||||  |||| :|||||| |||||||||:|||:|| |
Db        104 LRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKV 163

Qy        753 KIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCL 812
              |||||| ||| |||||||||||||||| | :|:| ||||||||||||:|||||||||
Db        164 TIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCL 223

Qy        813 LDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGL 872
              ||::||::  :||| |||||:||||||:|||:  |||||||||||||:|  ||||||||
Db        224 LDYIREHKDNIGSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGL 283
```

```
Qy        873 ARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 932
Db        284 AKQLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGI 343

Qy        933 PAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRF 992
Db        344 PASEISSVLEKGERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRY 403

Qy        993 VVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHH 1051
Db        404 LVIQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF-------------- 449

Qy       1052 RHRSSSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQS 1106
Db        450 --NSPST-----------------SRTPLLSSLSATSNNSATNCIDRNGG--------- 480

Qy       1107 LPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAA 1166
Db        481 ---H----------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQIQIY 513

Qy       1167 RPAGAT-LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203
Db        514 NYISLTAISKLPMDSRYQN-------SHSTAVDNPEYL 544


RESULT 15
Q9WVF5
ID   Q9WVF5      PRELIMINARY;      PRT;   655 AA.
AC   Q9WVF5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor (Epidermal growth factor receptor
DE   isoform 3).
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT   Carboxy-Terminal Truncated Receptors.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative Egfr transcripts encoding truncated receptor
RT   isoforms.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AF124513; AAD44149.1; -.
DR   EMBL; AF275366; AAG28047.1; -.
DR   EMBL; AF275364; AAG28047.1; JOINED.
DR   EMBL; AF275365; AAG28047.1; JOINED.
DR   EMBL; AK004944; BAB23688.1; -.
DR   EMBL; AK004883; BAB23641.1; -.
DR   EMBL; AK004911; BAB23662.1; -.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR002174; Furin-like.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 3.
KW   Receptor.
SQ   SEQUENCE   655 AA;  72906 MW;  6B34063B1BC928CB CRC64;

  Query Match           21.5%;  Score 1468.5;  DB 11;  Length 655;
  Best Local Similarity   43.9%;  Pred. No. 1.9e-102;
  Matches  286;  Conservative   90;  Mismatches  234;  Indels   41;  Gaps   11;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPF 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 -----NNFTVSFWLRVPKVSASHLEPDEEGACQPCPINCTHSCVDLDDKGC 648
Db        598 GIMGENNTLV--WKYA----------DANNVCHLCHANCTYGCAGPGLQGC 636
```

Search completed: July 22, 2003, 09:01:39
Job time : 54.5887 secs

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-579-593-14
Perfect score: 6841
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6627	96.9	1255	21	AAY92620	Human heregulin 2
2	6627	96.9	1255	22	AAE12130	Human tyrosine kin
3	6627	96.9	1255	22	AAB60167	HER2 transgene pla
4	6627	96.9	1255	23	AAU74545	Human HER2 (ErbB2)
5	6621	96.8	1255	17	AAW01111	HER-2/neu protein.
6	6621	96.8	1255	20	AAW92406	Human HER-2/neu on
7	6621	96.8	1255	21	AAB21198	Human HER-2/neu pr
8	6621	96.8	1255	21	AAY84780	Amino acid sequenc
9	6621	96.8	1255	22	AAB85458	Human HER-2/neu pr
10	6621	96.8	1255	22	AAG88267	HER2/neu amino aci
11	6621	96.8	1255	23	AAE24067	Human Her-2 protei
12	6621	96.8	1255	23	AAB20479	Human Her-2/neu pr
13	6621	96.8	1255	23	AAM51143	Human Her-2/neu on
14	6621	96.8	1255	23	AAU77114	Human Her-2/neu po
15	6578	96.2	1433	14	AAR39568	Sequence of c-erbB
16	6457	94.4	1223	23	AAU98923	Human breast cance
17	6304	92.2	1200	21	AAB21208	Human HER-2/neu pr
18	5837.5	85.3	1256	21	AAB21199	Rat HER-2/neu prot
19	5837.5	85.3	1256	23	AAM51144	Rat Her-2/neu onco
20	5810.5	84.9	1256	21	AAB21206	Mouse Her-2/neu pr
21	5810.5	84.9	1256	22	AAG62860	Amino acid sequenc
22	5810.5	84.9	1256	23	AAM51151	Mouse Her-2/neu on
23	4707	68.8	919	21	AAB21203	Human HER-2/neu fu
24	4707	68.8	919	23	AAM51148	Her-2/neu extracel
25	3960.5	57.9	920	23	AAM51152	Mouse Her-2/neu ex
26	3960.5	57.9	926	23	AAM51153	Mouse Her-2/neu ex
27	3591	52.5	712	21	AAB21204	Human HER-2/neu fu
28	3591	52.5	712	23	AAM51149	Her-2/neu extracel
29	3445	50.4	782	18	AAW19764	Her2-GM-CSF immuno
30	3443	50.3	653	21	AAB21200	Extracellular HER-
31	3443	50.3	653	23	AAM51145	Human Her-2/neu on
32	3405	49.8	645	22	AAB60408	Human ErbB2 oncopr
33	3405	49.8	645	22	AAB61593	Human ErbB2 extrac
34	3340	48.8	951	21	AAY44993	DC8scFv-erbB2EC fu
35	3237	47.3	624	11	AAR08222	Extracellular port
36	3110	45.5	1210	21	AAB19259	Amino acid sequenc
37	3110	45.5	1210	21	AAY50616	Human EGF receptor
38	3110	45.5	1210	23	AAE23019	Human Her-1 protei
39	3110	45.5	1210	23	AAM50768	Human epidermal gr
40	3108	45.4	1210	22	AAB68420	Amino acid sequenc
41	3084	45.1	583	23	AAE20483	Human protein for
42	3084	45.1	587	23	AAE20481	Human protein for
43	3083	45.1	589	23	AAE20484	Human protein for
44	3083	45.1	600	23	AAE20482	Human protein for
45	3069	44.9	1210	23	ABP51768	Human epidermal gr

ALIGNMENTS

RESULT 1
AAY92620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

```
FT                   /label= Cysteine_rich_domain
FT   Region          210..224
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          250..264
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          324..483
FT                   /label= Ligand_binding_domain
FT   Region          325..339
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          369..383
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          465..479
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          484..623
FT                   /label= Cysteine_rich_domain
FT   Region          579..593
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          624..654
FT                   /label= Transmembrane_domain
FT   Region          632..652
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          653..667
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          655..1010
FT                   /label= Tyrosine_kinase_domain
FT   Region          661..675
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          695..709
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          710..730
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          1011..1235
FT                   /label= C-terminal_domain
XX
PN   WO200020027-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;    99WO-DK00525.
XX
PR   05-OCT-1998;    98DK-0001261.
PR   20-OCT-1998;    98US-0105011.
XX
PA   (MEBI-) M & E BIOTECH AS.
XX
PI   Steinaa L,  Mouritsen S,  Nielsen KG,  Haaning J,  Leach D,  Dalum I;
PI   Gautam A,  Birk P,  Karlsson G;
XX
DR   WPI; 2000-349917/30.
DR   N-PSDB; AAA09455.
XX
PT   Inducing immune responses to weakly immunogenic, tumor associated
PT   peptide antigens for the treatment of breast and prostate cancer
XX
PS   Claim 62; Page 193-198; 220pp; English.
XX
CC   This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC   Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC   response.  Subdominant CTL epitopes, antibody binding regions and
CC   cysteine residues involved in disulfide bonds are preserved in the
CC   immunogenized forms.  Regions suitable for the insertion of foreign T
CC   helper epitopes were identified (see features table).  The method
CC   is used for inducing immune responses against weakly immunogenic
CC   cell-associated peptide antigens (PA) such as those associated with
CC   cancers (self-proteins), e.g. human prostate specific membrane antigen
CC   (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC   The method comprises effecting simultaneous presentation by antigen
CC   producing cells (APCs) of the animals immune system of: (1) at least 1
CC   CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC   B-cell group derived from the cell-associated PA; and (2) at least 1
CC   first T helper cell group which is foreign to the animal.  Analogues of
CC   human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC   part of all known and predicted CTL and B-cell epitopes of the respective
CC   PA and including at least one foreign T helper epitope are also claimed.
CC   The method is used to treat prostate, prostate/breast or breast cancer
CC   when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6627;  DB 21;  Length 1255;
  Best Local Similarity   97.0%;  Pred. No. 0;
  Matches 1227;  Conservative    6;  Mismatches   18;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |               |   ::    ||||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770
```

```
Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Region   774..782
FT   /note= "Antigenic epitope"
XX
PN WO200168677-A2.
XX
PD 20-SEP-2001.
XX
PF 16-MAR-2001; 2001WO-US40328.
XX
PR 16-MAR-2000; 2000US-0527487.
XX
PA (GENZ) GENZYME CORP.
XX
PI Nicolette CA;
XX
DR WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT -
XX
PS Claim 4; Page 63-67; 69pp; English.
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6627; DB 22; Length 1255;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::::::::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |         |    |::     ||||||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650
```

```
Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX
DR WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6627; DB 22; Length 1255;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |               |   |::     ||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoelic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
PI Erickson S, Schwall R, Sliwkowski M;
XX
DR WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoelic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6627; DB 23; Length 1255;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
```

```
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX
AC AAW01111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Domain   676..1255
FT   /label= Intracellular_domain
FT   /note= "claimed domain, useful for immunisation"
XX
PN WO9630514-A1.
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.
XX
CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neo oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 96.8%; Score 6621; DB 17; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
```

```
Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |              |   |::    ||||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

```
RESULT 6
AAW92406
ID   AAW92406 standard; Protein; 1255 AA.
XX
AC   AAW92406;
XX
DT   21-APR-1999  (first entry)
XX
DE   Human HER-2/neu oncogene protein.
XX
KW   HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW   malignancy; treatment; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          676..1255
FT                   /note= "region which elicits immune response"
XX
PN   US5869445-A.
XX
PD   09-FEB-1999.
XX
PF   01-APR-1996;   96US-0625101.
XX
PR   01-APR-1996;   96US-0625101.
PR   17-MAR-1993;   93US-0033644.
PR   12-AUG-1993;   93US-0106112.
PR   31-MAR-1995;   95US-0414417.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Disis ML;
XX
DR   WPI; 1999-152835/13.
DR   N-PSDB; AAX01912.
XX
PT   Use of HER-2/neu polypeptides - for eliciting an immune response to
PT   an HER-2/neu associated malignancy, particularly for treating or
PT   preventing tumours
XX
PS   Claim 3; Column 31-38; 26pp; English.
XX
CC   This sequence represents the human HER-2/neu oncogene protein. A fragment
CC   of this protein is used in a method for eliciting or enhancing an immune
CC   response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC   B cells to produce an immune response to the HER-2/neu protein. The
CC   method can be used for immunisation against a malignancy in which the
CC   HER-2/neu oncogene is associated and in the treatment of an existing
CC   tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ   Sequence   1255 AA;
```

Query Match 96.8%; Score 6621; DB 20; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
```

```
Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |          |   |::     ||||||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN WO200044899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX
PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;

Query Match 96.8%; Score 6621; DB 21; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
```

```
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716

Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776

Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836

Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896

Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956

Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016

Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076

Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136

Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196

Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256

Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261

Db       1251 LDVPV 1255


RESULT 8
AAY84780
ID   AAY84780 standard; Protein; 1255 AA.
XX
AC   AAY84780;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW   SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW   tumor cell proliferation; tissue degeneration; arthropathy;
KW   bone resorption; inflammatory disease; degenerative disorder;
KW   wound healing.
XX
OS   Homo sapiens.
XX
PN   WO200020579-A1.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-CA00912.
XX
PR   02-OCT-1998;   98US-0165192.
XX
PA   (UYMC-) UNIV MCMASTER.
XX
PI   Muller WJ,  Siegel PM;
XX
DR   WPI; 2000-303768/26.
DR   N-PSDB; AAA14812.
XX
PT   Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT   erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS   Claim 3; Fig 2; 60pp; English.
XX
CC   The present sequence represents a SPLICE erbB-2 receptor protein. The
CC   protein has an in-frame deletion of 16 amino acids, 2 of which are
CC   conserved cysteine residues, compared to the unspliced protein. The
CC   erbB-2 polynucleotide is used to construct probes for detecting
CC   disorders of cell transformation such as cancer. Antibodies to the
CC   protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC   (e.g. antisense oligonucleotides) which inhibit the expression of
CC   SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC   treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC   for treating conditions involving damaged cells including conditions
CC   in which degeneration of tissue occurs, such as arthropathy, bone
CC   resorption, inflammatory diseases, degenerative disorders of the
CC   central nervous system and wound healing.
XX
SQ   Sequence   1255 AA;

  Query Match             96.8%;  Score 6621;  DB 21;  Length 1255;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1225;  Conservative    7;  Mismatches   19;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
```

```
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255


RESULT 9
AAB85458
ID   AAB85458 standard; Protein; 1255 AA.
XX
AC   AAB85458;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human HER-2/neu protein.
XX
KW   Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW   oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS   Homo sapiens.
XX
PN   WO200153463-A2.
XX
PD   26-JUL-2001.
XX
PF   19-JAN-2001; 2001WO-US01850.
XX
PR   21-JAN-2000; 2000US-0177545.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Cheever MA,  Hand-Zimmermann S;
XX
DR   WPI; 2001-476112/51.
DR   N-PSDB; AAH23392.
XX
PT   New antigen-presenting cells, useful as vaccines for eliciting or
PT   enhancing an immune response to HER-2/neu protein, particularly useful
PT   for treating or preventing cancer, e.g. breast cancer  -
XX
PS   Claim 2; Page 41-46; 49pp; English.
XX
CC   The invention provides an isolated antigen-presenting cell, which
CC   expresses at least an immunogenic portion of a polypeptide that produces
CC   an immune response to HER-2/neu protein. The antigen-presenting cells are
CC   useful as vaccines for eliciting or enhancing an immune response to
CC   HER-2/neu protein, particularly in treating or preventing malignancies in
CC   which the HER-2/neu oncogene is associated. Specifically, these are
CC   useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC   colon, lung or prostate cancers. The present sequence represents
CC   the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ   Sequence   1255 AA;

  Query Match             96.8%;  Score 6621;  DB 22;  Length 1255;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1225;  Conservative    7;  Mismatches   19;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |||||||||||||||||||||||    |||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

```
RESULT 10
AAG88267
ID   AAG88267 standard; Protein; 1255 AA.
XX
AC   AAG88267;
XX
DT   11-SEP-2001  (first entry)
XX
DE   HER2/neu amino acid sequence.
XX
KW   Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW   immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW   tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS   Homo sapiens.
XX
PN   WO200141787-A1.
XX
PD   14-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US33591.
XX
PR   10-DEC-1999;   99US-0458299.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
PI   Keogh E;
XX
DR   WPI; 2001-374995/39.
XX
PT   An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT   cellular immune responses for the prevention and treatment of cancer -
XX
PS   Disclosure; Page 15; 199pp; English.
XX
CC   The present invention describes isolated prepared HER2/neu epitopes (I).
CC   Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC   culture in vitro and binds to a complex of an epitope (I), bound to a
CC   human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC   and a second epitope and the peptide is less than 50 contiguous amino
CC   acids that have 100% identity with a native peptide sequence of HER2/neu;
CC   (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC   excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC   (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC   and immunostimulant activities, and can be used in vaccines. (I), (II)
CC   and (III) are useful for inducing cellular immune responses for the
CC   prevention and treatment of cancer. (I) and (II) are useful for
CC   monitoring or evaluating an immune response to a tumour-associated
CC   antigen when incubated with a T lymphocyte sample form a patient and
CC   detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC   based vaccines mean that immunosuppressive epitopes that may be present
CC   in whole antigens may be avoided. Selected epitopes may be combined to
CC   enhance immunogenicity. The possible pathological side effects caused by
CC   infectious agents or whole protein antigen is eliminated. The vaccine
CC   provides the ability to direct and focus an immune response to multiple
CC   selected antigens from the same pathogen. Epitope-based anti-tumour
CC   vaccines provides the opportunity to combine epitopes derived from
CC   multiple tumour-associated molecules addressing the problem of tumour-
CC   tumour variability and reducing the likelihood of tumour escape due to
CC   antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match           96.8%;  Score 6621;  DB 22;  Length 1255;
  Best Local Similarity 96.8%;  Pred. No. 0;
  Matches 1225;  Conservative    7;  Mismatches   19;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |                |   |::     ||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowsert LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;

```
  Query Match             96.8%;  Score 6621;  DB 23;  Length 1255;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1225;  Conservative    7;  Mismatches   19;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
```

```
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |           |   |::     ||||||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255
```

```
RESULT 12
AAE20479
ID   AAE20479 standard; Protein; 1255 AA.
XX
AC   AAE20479;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Human Her-2/neu protein.
XX
KW   Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW   human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1021..1030
FT                   /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN   WO200214503-A2.
XX
PD   21-FEB-2002.
XX
PF   14-AUG-2001; 2001WO-US41733.
XX
PR   14-AUG-2000; 2000US-225152P.
PR   28-SEP-2000; 2000US-236428P.
PR   21-FEB-2001; 2001US-270520P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Hand-zimmermann S,  Cheever MA,  Foy TM,  Lodes MJ,  Kalos MD;
PI   Mcneill PD,  Vedvick TS;
XX
DR   WPI; 2002-280758/32.
DR   N-PSDB; AAD32743.
XX
PT   Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT   prevention and diagnosis of cancer, preferably breast cancer  -
XX
PS   Disclosure; Page 114-117; 129pp; English.
XX
CC   The invention relates to an isolated Her-2/Neu polypeptide composition
CC   effective for eliciting an immune response. The invention is useful for
CC   eliciting an immune response in a patient, where the patient is human
CC   leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC   The composition is useful for the therapy and diagnosis of cancer,
CC   preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC   and other compositions for the diagnosis, prevention and treatment of
CC   human malignancies, for stimulating and/or expanding T cells specific for
CC   Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC   patient. The invention is useful for stimulating a T cell response in a
CC   human patient, as probe or primer for nucleic acid hybridisation, to
CC   selectively form duplex molecules with complementary stretches of the
CC   entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC   length gene from a suitable library, and to direct expression of a
CC   polypeptide in appropriate host cells. The composition is useful in
CC   prophylactic or therapeutic applications and for the treatment of cancer,
CC   preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC   associated malignancies. The invention is useful in gene therapy. The
CC   present sequence is human Her-2/neu protein.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match             96.8%;  Score 6621;  DB 23;  Length 1255;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1225;  Conservative    7;  Mismatches   19;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
              |         |   |::    ||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
              |||||
Db       1251 LDVPV 1255


RESULT 13
AAM51143
ID   AAM51143 standard; Protein; 1255 AA.
XX
AC   AAM51143;
XX
DT   17-JUN-2002  (first entry)
XX
DE   Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX
KW   Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW   tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..653
FT                   /note= "extracellular domain"
FT   Domain          676..1255
FT                   /note= "intracellular domain"
FT   Domain          990..1255
FT                   /note= "phosphorylation domain"
XX
PN   WO200212341-A2.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-US24283.
XX
PR   03-AUG-2000; 2000US-0632507.
XX
PA   (CORI-) CORIXA CORP.
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Cheever MA,  Gheysen D;
XX
DR   WPI; 2002-241743/29.
DR   N-PSDB; ABA92250.
XX
PT   Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT   or enhancing an immune response to the protein, has Her-2/neu
PT   extracellular domain fused to Her-2/neu intracellular or
PT   phosphorylation domain  -
XX
PS   Claim 68; Fig 7; 141pp; English.
XX
CC   The present sequence is that of human Her-2/neu (p185 glycoprotein
CC   or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC   vaccines.  The Her-2/neu gene is amplified and p185 is overexpressed
CC   in a variety of cancers, including breast, ovarian, colon, lung and
CC   prostate cancer.  Her-2/neu is a member of the tyrosine kinase
CC   family of receptor-like glycoproteins.  It comprises an extracellular
CC   domain with homology to the epidermal growth factor receptor
CC   (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC   intracellular domain that also shows homology to EGFR.  Its
CC   overexpression correlates with a poor prognosis in breast and
CC   ovarian cancers.  The invention provides Her-2/neu fusion
CC   proteins, nucleic acids encoding them, viral vectors, and vaccines
CC   comprising the fusion proteins or nucleic acid molecules.  In
CC   preferred fusion proteins, the extracellular domain of a Her-2/neu
CC   protein is fused to a Her-2/neu intracellular domain or
CC   phosphorylation domain (or its DeltaPD fragment).  An immune
CC   response to Her-2/neu protein is elicited or enhanced by
CC   administering the fusion protein in the form of a vaccine, or by
CC   transfecting cells of an animal ex vivo with a nucleic acid
CC   encoding the fusion protein, and delivering the transfected cells
CC   to the animal.  The fusion proteins, nucleic acids, and isolated
CC   specific T-cells are useful for inhibiting the development of a
CC   cancer, especially breast, ovarian, colon, lung or prostate cancer
CC   in a patient.  T cells that specifically react with a Her-2/neu
CC   fusion protein can be used to remove tumour cells from a sample in
CC   order to inhibit the development of cancer in a patient.
XX
SQ   Sequence   1255 AA;

  Query Match             96.8%;  Score 6621;  DB 23;  Length 1255;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1225;  Conservative    7;  Mismatches   19;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
```

```
Qy          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy         361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy         601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
               |          |   |::     ||||||||||||||||||||||||||||||
Db         601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy         657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
               ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy         717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy         777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy         837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy         897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy         957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy        1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy        1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy        1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy        1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy        1257 LDVPV 1261
               |||||
Db        1251 LDVPV 1255
```

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide -
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 96.8%; Score 6621; DB 23; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
                 |          |   |::     ||||||||||||||||||||||||||||||||
Db        601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy        657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
                 ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy        717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy       1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy       1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
                 |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy       1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy       1257 LDVPV 1261
                 |||||
Db       1251 LDVPV 1255
```

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells. such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isolelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

```
  Query Match             96.2%;  Score 6578;  DB 14;  Length 1433;
  Best Local Similarity   96.3%;  Pred. No. 0;
  Matches 1218;  Conservative    9;  Mismatches   24;  Indels   14;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQVCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEHCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
```

```
Qy       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
Db       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy       361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
Db       361 IQEFAGCKRIFGSLAFLPESFDGDPASNTAPLQPEHLQVFETLEQITGYLYISAWPDSLP 420

Qy       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
Db       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLSFVHTV 480

Qy       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
Db       541 VEECRVLQGLPREYVNASHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy       601 PFNNFTVSFWLRVPKVSASHLE----PDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 656
Db       601 PSG----------VKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP 650

Qy       657 LTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
Db       651 LTSIISAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Qy       717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
Db       711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770

Qy       777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
Db       771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830

Qy       837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
Db       831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890

Qy       897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
Db       891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950

Qy       957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
Db       951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010

Qy      1017 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1076
Db      1011 DDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEA 1070

Qy      1077 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db      1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130

Qy      1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196
Db      1131 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGA 1190

Qy      1197 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db      1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250

Qy      1257 LDVPV 1261
Db      1251 LDVPV 1255
```

Search completed: July 22, 2003, 08:41:51
Job time : 44.1589 secs

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-632-652-14
Perfect score: 6834
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6623	96.9	1255	1	A24571	protein-tyrosine k
2	5864	85.8	1260	1	TVRTNU	protein-tyrosine k
3	5841.5	85.5	1254	2	I48161	p-185 precursor -
4	3113	45.6	1210	1	GQHUE	epidermal growth f
5	3080	45.1	1210	2	A53183	epidermal growth f
6	3049.5	44.6	1223	1	TVCHLV	epidermal growth f
7	2920.5	42.7	1308	2	A47253	epidermal growth f
8	2627	38.4	1166	1	S06142	protein-tyrosine k
9	2389.5	35.0	1342	2	A36223	kinase-related tra
10	2302.5	33.7	1339	2	JC4387	epidermal growth f
11	1721.5	25.2	698	1	TVFVLV	protein-tyrosine k
12	1658	24.3	604	1	TVYUH	protein-tyrosine k
13	1611.5	23.6	1330	1	GQFFE	epidermal growth f
14	1602	23.4	544	2	S35745	protein-tyrosine k
15	1595	23.3	545	2	S00727	kinase-related tra
16	1578	23.1	540	2	B44776	protein-tyrosine k
17	1576	23.1	540	1	TVFVEB	protein-tyrosine k
18	1509	22.1	644	2	A36325	epidermal growth f
19	1274	18.6	1323	2	E88257	protein let-23 [im
20	1274	18.6	1374	2	S70712	protein-tyrosine k
21	1195	17.5	1369	2	S70713	protein-tyrosine k
22	1150	16.8	1717	1	A45558	epidermal growth f
23	1126	16.5	527	2	A42032	epidermal growth f
24	975.5	14.3	843	2	A27131	epidermal growth f
25	806.5	11.8	346	2	S13807	protein-tyrosine k
26	754.5	11.0	311	2	S13808	protein-tyrosine k
27	713	10.4	1363	2	T43220	insulin-like growt
28	694	10.2	1382	1	INHUR	insulin receptor p
29	693.5	10.1	1372	2	A34157	insulin receptor p
30	690	10.1	1383	2	A36080	insulin receptor p
31	680	10.0	1477	2	T18534	protein-tyrosine k
32	680	10.0	1607	2	T43212	insulin-like growt
33	679.5	9.9	1300	2	A36502	insulin receptor-r
34	669	9.8	1268	2	B36502	insulin receptor-r
35	644	9.4	1367	1	IGHUR1	insulin-like growt
36	629	9.2	1371	2	A33837	insulin-like growt
37	622.5	9.1	2148	1	A56081	insulin receptor -
38	617	9.0	1390	2	T30346	insulin receptor -
39	616	9.0	2101	2	S57245	insulin receptor (
40	591	8.6	1114	1	S05582	protein-tyrosine k
41	589.5	8.6	1028	2	A45388	protein-tyrosine k
42	586	8.6	1091	2	S33596	protein-tyrosine k
43	579	8.5	1052	2	A46166	protein-tyrosine k
44	578.5	8.5	952	2	I50612	protein-tyrosine k
45	576.5	8.4	984	2	A39753	protein-tyrosine k

ALIGNMENTS

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito,
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal gro
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/ep
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seebu
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares ch
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517,'RALL',522,'S',524-654,'V',656-1169,'A',1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:g183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcript
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

```
A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TMM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

  Query Match             96.9%;  Score 6623;  DB 1;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 3.6e-265;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :

Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                 : |  |    | ||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related prote
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Co
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
```

```
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

  Query Match             85.8%;  Score 5864;  DB 1;  Length 1260;
  Best Local Similarity   86.2%;  Pred. No. 5.7e-234;
  Matches 1084;  Conservative   48;  Mismatches  123;  Indels    2;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||||
Db          4 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 63

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| |||||||||||||||:|||||::|||||||||||||||| ||||||||
Db         64 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 123

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||||: |||||||| :||||:| |||||
Db        124 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 183

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |   |||||||| ||:| || : ||||| |||| || |:|  |||||| ||||||||
Db        184 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 243

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |
Db        244 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 303

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||||||||::::
Db        304 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 363

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
              || |||||||||||||||||:|  |||:|||||||||||||||||||||||||||||||
Db        364 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 423

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              ||||||||::|||||||:|||||||||||||||||||||||| | |||||||
Db        424 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 483

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||: ||||:: ||  || |: ||| |||||||||||||| |||||
Db        484 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 543

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : |||||||||||| | |||| |||||| |||||   |||
Db        544 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 603

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNF 658
              |||||||||||||||||:||||| |||||||||||||||::|||||||||:| |
Db        604 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRASPVTFIIATV 663

Qy        659 TVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 718
                    :  |    | |||:||||||||||||||||||||||||||||||||||||
Db        664 EGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 723

Qy        719 ELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVG 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 ELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVG 783

Qy        779 SPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLED 838
              ||||||||||||||||||||||||||||||:||||||||||:||||||||||
Db        784 SPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYLED 843

Qy        839 VRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRR 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 VRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRR 903

Qy        899 FTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKC 958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 FTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKC 963

Qy        959 WMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLV 1018
              |||||||||||||||||||||||||||||||||||||:||:|||||||||||||||||||
Db        964 WMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDDDMGDLV 1023

Qy       1019 DAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPS 1078
              |||||||||||||| ||| || |   | ||||||||||||:|||||||||| ||||||||
Db       1024 DAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPPRSPLAPS 1083

Qy       1079 EGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPE 1138
              |||||||||||| ||  ||||||  || ||||||||||:||| ||||||||| ||||||
Db       1084 EGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPLACSPQPE 1143

Qy       1139 YVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLT 1198
              |||| :|:|||| |||||  |||||||| ||||||||||||||||||||||||||||||
Db       1144 YVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLV 1203

Qy       1199 PQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |: | |  ||| |||||||||||||||:  |:| ||| |:|||||||||||||||||
Db       1204 PREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGLDVPV 1260


RESULT 3
I48161
p-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishi
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MUID:94193007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

  Query Match             85.5%;  Score 5841.5;  DB 2;  Length 1254;
  Best Local Similarity   85.7%;  Pred. No. 4.7e-233;
  Matches 1076;  Conservative   57;  Mismatches  121;  Indels    1;  Gaps    1;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| | |||||||| ||| |||||||||||||||||||||||::|||||||||||||
Db          1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:||||:||| |||||||||||||||| ||||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||:| |  || :| |||||||||||||||||||: |||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                 ||||||||| ||:| || : ||| | ||||:|| |:        :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||:|||||||||:|||||||||||||||||| ||||||||:::::
Db        301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360
```

```
Qy        361 ELEPAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||:|:|  ||| ||||||||||||||||||||||||||||
Db        361 IQEPAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||:|||||:||:|||||| |||||| |||||||||||||| ||| ||||||||||
Db        421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||: |  |:||  :  ||: ||| |||||||||||||||| |||||||
Db        481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |:|||| :|||||||| :|||||||||||||| : || | ||||| || |||| |||||||
Db        541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||:|||| |||||||||||||||::|||||||| |||
Db        601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              |  |    | |||:||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGLGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||:|||||||||||||:|||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              |||||||||| |||||||    | ||||||||||||:||||:||| || ||||||||||
Db       1021 EEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||:|:|||||  || ||:   | ||||||||||:|||:|||||||| ||||||||
Db       1081 AGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLACSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||:||||||    |||||  |||||||| |||||||||||||| |||||||||  |:
Db       1141 NQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVENPEYLVPR 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              ||:| ||| |||  |||||||||||| |||:||:||:|||||||||||||||
Db       1201 GGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDVPV 1254


RESULT 4
GQHUE
epidermal growth factor receptor precursor - human
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Homo sapiens (man)
C;Date: 15-Nov-1984 #sequence_revision 27-Nov-1985 #text_change 11-Jun-1999
C;Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33
R;Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y
rg, P.H.
Nature 309, 418-425, 1984
A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210 <ULL>
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epiderma
A;Reference number: A25772; MUID:85270438; PMID:2991899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identificat
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript ter
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 'RCAWRRA',150-187,'KSVIQAV',195,'M',197,'A',199-222,'S',224-304,'RA',30
',798-799,'TD',802-811,'R',813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF
R;Lin, C.R.; Chen, W.S.; Kruiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verm
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30,'S',32-51;454-467 <WEB>
R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A;Title: Identification of residues in the nucleotide binding site of the epidermal
A;Reference number: A60143; MUID:85182650; PMID:2985580
A;Accession: A60143
A;Molecule type: protein
A;Residues: 740-744,'X',746-747 <RUS>
R;Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and su
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity
```

```
A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain r
A;Reference number: A33331; MUID:90003233; PMID:2790960
A;Contents: annotation; internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:120610; OMIM:131550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1210/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <EE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <EE2>
F;646-668/Domain: transmembrane #status predicted <TMM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-726/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F;745/Active site: Lys #status experimental

  Query Match             45.6%;  Score 3113;  DB 1;  Length 1210;
  Best Local Similarity   49.5%;  Pred. No. 6.3e-121;
  Matches  625;  Conservative  173;  Mismatches  363;  Indels  102;  Gaps   21;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
Db        127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy        189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
Db        184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy        248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
Db        244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy        308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
Db        304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy        367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
Db        362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy        427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
Db        422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy        487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
Db        482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCKL 541

Qy        547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606
Db        542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy        607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSFWLRV 666

Db        602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLLL 658

Qy        667 PKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVL 726
Db        659 VVALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVL 718

Qy        727 GSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLL 786
Db        719 GSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLL 778

Qy        787 GICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDL 846
Db        779 GICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDL 838

Qy        847 AARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVW 906
Db        839 AARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVW 898

Qy        907 SYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECR 966
Db        899 SYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSR 958

Qy        967 PRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLV 1025
Db        959 PKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLI 1018

Qy       1026 PQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDV 1085
Db       1019 PQQGFF----------------------------------SSPSTSRTPLLSSLSATSN- 1043

Qy       1086 FDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQP 1143
Db       1044 -NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------PVPEYINQ- 1095

Qy       1144 DVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-TPQGG 1202
Db       1096 SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNPEYLNTVQ-- 1143

Qy       1203 AAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTPTAENP 1246
Db       1144 -------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGS-TAENA 1195

Qy       1247 EYL 1249
Db       1196 EYL 1198


RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C;Accession: A53183; A43818; S24942; A28941; S45325; I49643
R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF recep
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:U03425
R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A;Title: Comparison of EGF receptor sequences as a guide to study the ligand bindin
A;Reference number: A43818; MUID:91232866; PMID:2030916
A;Accession: A43818
A;Molecule type: mRNA
A;Residues: 1-714 <AVI>
A;Cross-references: GB:X59698
R;Eisinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A;Reference number: S24942
A;Accession: S24942
```

```
A;Molecule type: mRNA
A;Residues: 969-971,'K',973-1115,'D' <EIS>
A;Cross-references: EMBL:Z12608
R;Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A;Reference number: A28941; MUID:88330814; PMID:3138233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971,'K',973-1210 <VER>
A;Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A;Reference number: I49643; MUID:93126380; PMID:7678348
A;Accession: I49643
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 12-20,22-132 <RES>
A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
C;Genetics:
A;Gene: EGFR
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphopro
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TMM>
F;712-977/Domain: protein kinase homology <KIN>
F;720-728/Region: protein kinase ATP-binding motif
F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

  Query Match             45.1%;  Score 3080;  DB 2;  Length 1210;
  Best Local Similarity   49.3%;  Pred. No. 1.4e-119;
  Matches  627;  Conservative  164;  Mismatches  371;  Indels  110;  Gaps   23;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSF 662
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 WLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRK 722
Db        657 LFIVVVALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKK 716

Qy        723 VKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYV 782
Db        717 IKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHV 776

Qy        783 SRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLV 842
Db        777 CRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLV 836

Qy        843 HRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQ 902
Db        837 HRDLAAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQ 896

Qy        903 SDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMID 962
Db        897 SDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVKCWMID 956

Qy        963 SECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAE 1021
Db        957 ADSRPKFRELILEFSQMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVVDAD 1016

Qy       1022 EYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGA 1081
Db       1017 EYLIPQQGFF----------------NSPST-----------------SRTPLLSSLSA 1042

Qy       1082 GSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEY 1139
Db       1043 TSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------PVPEY 1092

Qy       1140 VNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-T 1198
Db       1093 VNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNPEYLNT 1141

Qy       1199 PQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTPT 1242
Db       1142 AQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIFKG-PT 1191

Qy       1243 AENPEYLGLDVP 1254
Db       1192 AENAEYLRVAPP 1203


RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Gallus gallus (chicken)
C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C;Accession: A27720; A00643
R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Ver
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in
A;Reference number: A27720; MUID:88261272; PMID:3260329
```

```
A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
pecific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TMM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

  Query Match             44.6%;  Score 3049.5;  DB 1;  Length 1223;
  Best Local Similarity   48.3%;  Pred. No. 2.6e-118;
  Matches  626;  Conservative  166;  Mismatches  360;  Indels  145;  Gaps   25;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |   :| ||  | |         :|| ||: ||      | |    |: :|  |:||  |||
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  |:||: |||| |||||| | |  :||: |:|:|| |:::::|||| |
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
               :| |          ||||| :: |:||| ||| |  ||:||  ||:|| ||   : :  |
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              |::|  :   :|  | | |    |||   ::||:||: :||  |: ||:| :|:|||| |
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: ||||||  |     |:  || || ||  |::   ||||:|:|||:||  |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||   |     ||  |:| ::|:||   |::|| |:|:  :|     |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  | || | :: || :| ||  :  | || |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
              ||  |:||::||||   :| ||| :  | |  ||||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: |  ||    |    || :::|  :  |  ||:  ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
                ||::| :||| |||:     ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIF 655
              ||  ||:||  :   : :||: |    || |  ||| |     :|||      |  ||
Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 NNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRIL 715
                        |   ||:  :|| |:|||||| |||||||||  |||| :|||
Db        657 AGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRIL 716

Qy        716 KETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 775
              |||| :|||||||||||||||:|||:|| |||||||| ||| ||||||||||||||||
Db        717 KETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMA 776

Qy        776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835
              | :|:| ||||||||||||:|||||||||::||::  :||| |||||:||||||:|
Db        777 SVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWCVQIAKGMNY 836

Qy        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895
              ||: |||||||||||||:| ||||||||||:||  || |||||||||||||||||||
Db        837 LEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPIKWMALESIL 896

Qy        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955
              | :|||||||||||||||||:|||||||| ||  :|||||||||||||||||||||
Db        897 HRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPICTIDVYMIM 956

Qy        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDM 1014
              |||||||:: ||:||||::|||:||||| |::||| :| :   || || ||||:|:|::||
Db        957 VKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFYRTLMEEEDM 1016

Qy       1015 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1074
              |:|||:||||| ||||                    | ||                 |:|
Db       1017 EDIVDADEYLVPHQGFF----------------NSPST-----------------SRTP 1042

Qy       1075 L-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGY 1127
              |     | |  : ::  |         :  |  |  : | :|||| |||     |:  ||:
Db       1043 LLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNFLEESIDDGF 1094

Qy       1128 VAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVF-- 1185
              :      | ||||||  : |: ||                    :  :| :  ::
Db       1095 L------PAPEYVNQ--LMPKKPS--------------------TAMVQNQIYNNISLT 1125

Qy       1186 -------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQ------- 1225
                          :  ||:|||||          :  |     |::  || |
Db       1126 AISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFESSPYWIQSGNHQIN 1177

Qy       1226 -DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1254
              | |:           |:      | |||||| :  |
Db       1178 LDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 1214


RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epider
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif
```

Query Match 42.7%; Score 2920.5; DB 2; Length 1308;
Best Local Similarity 44.6%; Pred. No. 5.5e-113;
Matches 604; Conservative 178; Mismatches 384; Indels 187; Gaps 26;

```
Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  |  ||||: ::|| |||:|  || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |        ||:|| |::||||| ||| : :|  ||| ||| |:||      |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |  | | |||| :   ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||   ||  ||:   |  :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||    ||  |:| : |: |:  | :.| |:|   : :          :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :   : ||:| || |: ||||:| | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||   :| ::   ||  : |  ||: |: ||:   ||||| ||::| :| ||: |:|
Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              | :  |  ||: |   |: | |:|:  ::| :|| ||  | |  |:|:|| | || :||
Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIF----NNF 658
              |::   |:  |:|: | :  | ||  ||| |           |:     :
Db        596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDC----------IYYPWTGHS 642

Qy        659 TVSFWLRVPKVSAS--------------HLEKRRQQKIRKYTMRRLLQETELVEPLTPS 703
              |:    | | ::|                :  ||:  :|  :|| | ||||||||||
Db        643 TLPQHARTPLIAAGVIGGLFILVIVGLTFAVYVRRKSIKKKRALRRFL-ETELVEPLTPS 701

Qy        704 GAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKA 763
              |  |||||:||||||||::|||||||||||||||||:|:|| ||||||||:| | | |||
Db        702 GTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKA 761

Qy        764 NKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLL 823
              | | :||| :|| :  |:: ||||:||: |:|||||||:||||::| |::  :||| ||
Db        762 NVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLL 821

Qy        824 NWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGK 883
              |||:|||||| |||: |||||||||||||||||||||||||||||: || ||:|||||
Db        822 NWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGK 881

Qy        884 VPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQ 943
              :|||||||| |  |:|||||||||||:|||||| |||||| |||||||||||||||||
Db        882 MPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQ 941

Qy        944 PPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDS 1002
              |||||||||:|||||||:: ||:|:|| :||||||||||::||| :| :   || ||
Db        942 PPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDS 1001

Qy       1003 TF"RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLG 1062
```

```
Db       1002 KFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIG 1049

Qy       1063 LEPSEEEAPRS----------------------------PLAP-SEGAGSDVFDGDLGMG 1093
                |     | |                            | || ::|| :::||     |
Db       1050 HSPPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNG 1109

Qy       1094 AAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVR 1146
                :  :   : |  |||| |||| |        : :||: |:   |: ||:|  :
Db       1110 TLRKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE-- 1167

Qy       1147 PQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQ 1206
                | |  : |              ||| ::       |::||||      |
Db       1168 -------ENPFVSRR--------------KNGDLQ-------ALDNPEYHNASNG---- 1194

Qy       1207 PHPPPA--------------------------------FSPAFDNLYYWDQDPPERGA 1232
                || |                                ||||  ||:   | |
Db       1195 --PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRST 1252

Qy       1233 --PPSTFKGTPT-------------AENPEYL 1249
                |   :  |              |||||||
Db       1253 LQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285
```

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transfo
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueler, W.; Raulf, F.; Telling, A.; Ro
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueler, W.; Schartl, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025,'N',1027-1098,'A',1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 38.4%; Score 2627; DB 1; Length 1166;
Best Local Similarity 44.7%; Pred. No. 5.7e-101;
Matches 568; Conservative 165; Mismatches 393; Indels 144; Gaps 31;

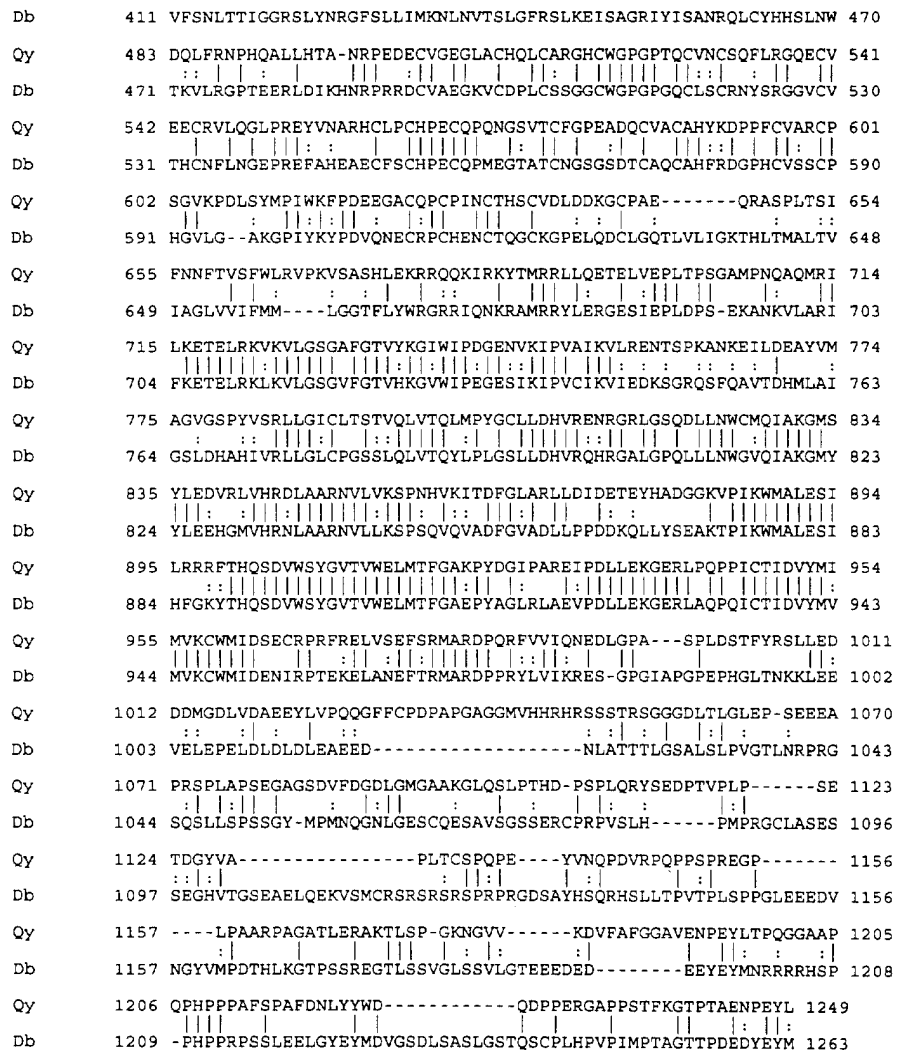
```
Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              ||| :  ||| |       ||    :|| ||  :: :    : |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||   |  ||||| |||| |||||  |:|  :||  ||::||  |:| |: | |: |
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                  | ::|       ||::||| :|||| ||| :  || ||  :|| | ||  | :
```

```
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
Db        417 NMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRY 476

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
Db        477 ANTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCW-PGPTMCVSCLHVDR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFC 596
Db        529 GGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKSAHFQDGPQC 588

Qy        597 VARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAE--QRASPLTSI 654
Db        589 IPRCPHGILGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGDIVSHSSLAVGL 647

Qy        655 FNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRI 714
Db        648 VSGLLITVIVALLIVV---LLRRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQAFLRI 703

Qy        715 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 774
Db        704 LKETEFKKDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQEVLDEAYVM 763

Qy        775 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMS 834
Db        764 ASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCVQIAKGMN 823

Qy        835 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI 894
Db        824 YLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPIKWMALESI 883

Qy        895 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Db        884 LQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPICTIEVYMI 943

Qy        955 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1014
Db        944 ILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSLSDRRLFSRLLSSDD- 999

Qy       1015 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1074
Db       1000 -DVVDADEYLLPYKRI----------NRQGS----------EP 1021

Qy       1075 LAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDGYVAPLTC 1133
Db       1022 CIPPTGH----------PVRENSITLRNISDPTQNALEKDLDGH------ 1055

Qy       1134 SPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKTLSPGKNGV 1180
Db       1056 ----EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYLNTNQNSL 1111

Qy       1181 VKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGT 1240
Db       1112 ---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGALTGNGMFL 1146

Qy       1241 PTAENPEYLG 1250
Db       1147 PAAENLEYLG 1156
```

RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C;Accession: A36223; I59164
R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epide
A;Reference number: A36223; MUID:90083234; PMID:2687875
A;Accession: A36223
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1342 <KRA>
A;Cross-references: GB:M29366
R;Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todarc
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A;Title: Molecular cloning and expression of another epidermal growth factor recept
A;Reference number: I59164; MUID:90311312; PMID:2164210
A;Accession: I59164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-559,'G',561-957,'F',959-1063,'G',1065-1342 <RES>
A;Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C;Genetics:
A;Gene: GDB:ERBB3; HER3
A;Cross-references: GDB:119880; OMIM:190151
A;Map position: 12q13-12q13
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; phosphotransferase
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

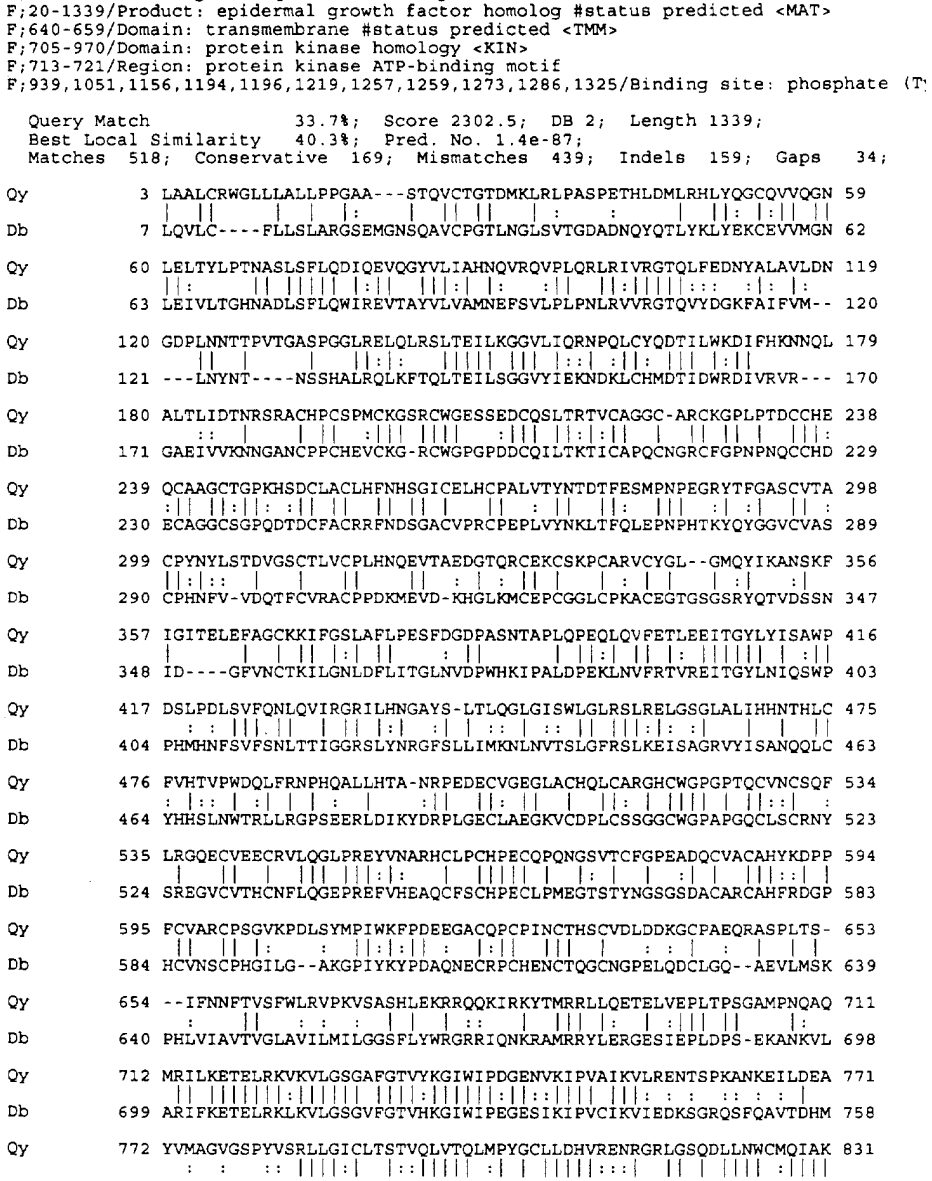
Query Match 35.0%; Score 2389.5; DB 2; Length 1342;
Best Local Similarity 40.0%; Pred. No. 3.8e-91;
Matches 526; Conservative 192; Mismatches 459; Indels 139; Gaps 33;

```
Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
Db        238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFQLEPNPHTKYQYGGVCVASCPHNFV-V 296

Qy        307 DVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF--IGITELE- 363
Db        297 DQTSCVRACPPDKMEVD-KNGLKMCEPCGGLCPKACEGTG-----SGSRFQTVDSSNIDG 350

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
Db        351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVFRTVREITGYLNIQSWPPHMHNFS 410

Qy        424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
```



RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res
C;Comment: This protein is a functional heregulin receptor that transduces signals
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase l
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prot
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TMM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (T

Query Match 33.7%; Score 2302.5; DB 2; Length 1339;
Best Local Similarity 40.3%; Pred. No. 1.4e-87;
Matches 518; Conservative 169; Mismatches 439; Indels 159; Gaps 34;



```
Db        759 LAVGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVKQHRETLGPQLLLNWGVQIAK 818

Qy        832 GMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMAL 891
              || |||:  :|||||| |||::|||: |:: |||:| ||  |: :      | |||||||
Db        819 GMYYLEEHSMVHRDLALRNVMLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKWMAL 878

Qy        892 ESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDV 951
              |||   ::|||||||||||||||||||||:|| |:   |||||||||||| || ||||||
Db        879 ESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPQICTIDV 938

Qy        952 YMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLED 1011
              ||:|||||||   || |:|| :||:|||||| |::||:    || :|      |:|
Db        939 YMVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAEPSVLTT 995

Qy       1012 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE--- 1068
               ::      |  | |:                              || | ||  ||
Db        996 KEL-----QEAELEPEL-----------------------------DLDLDLEAEEEGLA 1021

Qy       1069 -----------------EAPRSPLAPSEG--------AGSDVFDGDLGMGAAKGLQSLPT 1103
                              :| |:|| |         |    |  :  |  :  : :
Db       1022 TSLGSALSLPTGTLTRPRGSQSLLSPSSGYMPMNQSSLGEACLDSAVLGGREQFSRPISL 1081

Qy       1104 HDPSPLQRYSEDPTVPLPSETDGYV----APL-----TC-------SPQPE----YVNQP 1143
              | | |  |       |    ::|:|    | |       |       ||:|     | :|
Db       1082 H-PIPRGR-------PASESSEGHVTGSEAELQEKVSVCRSRSRSRSPRPRGDSAYHSQR 1133

Qy       1144 DVRPQPPSPREGP-----------LPAARPAGATLERAKTLSP-GKNGVV------KDVF 1185
                | :|   |            :|     ||:  |  |||  | : |:      :|
Db       1134 HSLLTPVTPLSPPGLEEEDGNGYVMPDTHLRGASSSREGTLSSVGLSSVLGTEEEDED-- 1191

Qy       1186 AFGGAVENPEYLTPQGGAAPQPHPP 1210
                 |  ||:  :   :| | ||
Db       1192 ------EEYEYMNRKRRGSP-PRPP 1209
```

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

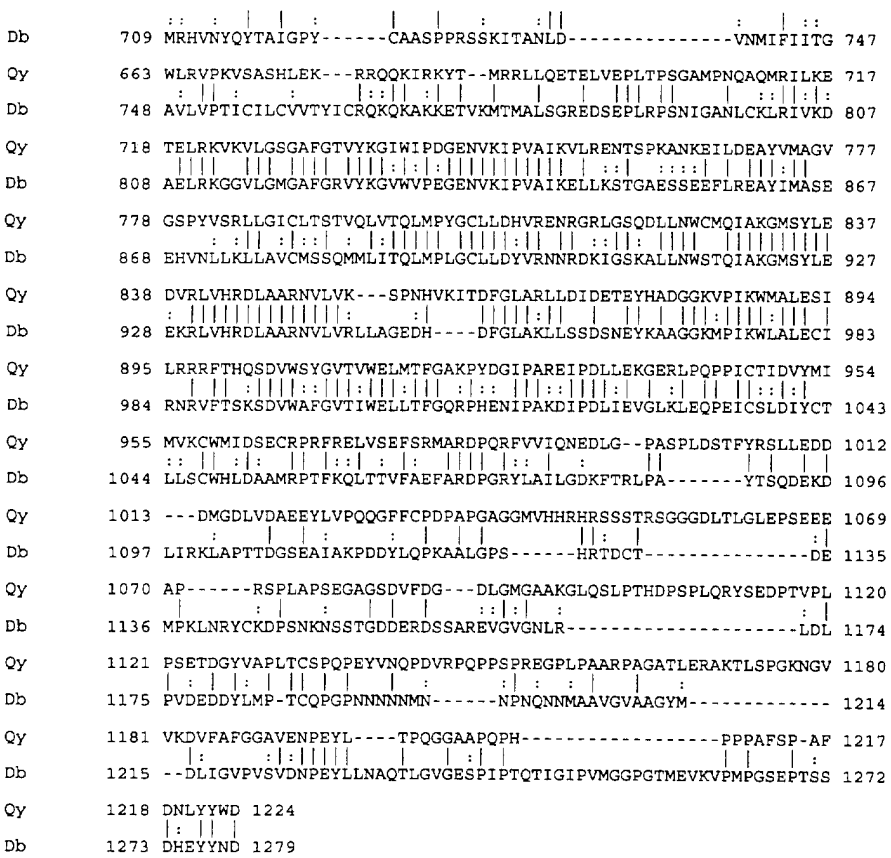
  Query Match   25.2%; Score 1721.5; DB 1; Length 698;
  Best Local Similarity   51.5%; Pred. No. 5.6e-64;
  Matches 368; Conservative   73; Mismatches 151; Indels 123; Gaps   16;

```
Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |: |||: | | ||  ||:||  :   : :||: |    || |  ||| |
Db         60 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 116

Qy        638 DDKGCPAEQRASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELV 697
               :|||     |  ||           |       |  ||:  :|| |:|||||| |||
```

```
Db        117 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELV 173

Qy        698 EPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRE 757
              |||||||   |||| :||||||| :|||||||||||||||:|||:|| |||||||| |||
Db        174 EPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELRE 233

Qy        758 NTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRL 817
              |||||||||||||||||| | :|:| |||||||||||||:|||||||||||::||::  :
Db        234 ATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNI 293

Qy        818 GSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEY 877
              ||| |||||:||||||:|||: ||||||||||||||:| |||||||||:||   || ||
Db        294 GSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEY 353

Qy        878 HADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 937
              ||:|||||||||||||| | :|||||||||||||||||||:|||||||| ||  :|||
Db        354 HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEK 413

Qy        938 GERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGP 996
              ||||||||||||||||||||||:: ||:||||::|||:|||||  |::||| :| :
Db        414 GERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHL 473

Qy        997 ASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGG 1056
              || || |||:|:|::|| |:|||:|||||  ||||                 | ||
Db        474 PSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF---------------NSPST---- 513

Qy       1057 GDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQR 1111
                                |:||     | |  : ::  |         :  |  |  : :||
Db        514 --------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQR 551

Qy       1112 YSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLER 1169
              || |||     |:  ||::     | ||||||  : |: ||
Db        552 YSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS------------------ 585

Qy       1170 AKTLSPGKNGVVKDVF---------------AFGGAVENPEYLTPQGGAAPQPHPPPAFS 1214
                 :  :| :  ::                 :   ||:|||||          :  |
Db        586 ---TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAK 634

Qy       1215 PAFDNLYYWDQ--------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1254
                |:  || |        | |:           |:     | |||||| :  |
Db        635 TVFESSPYWIQSGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 689
```

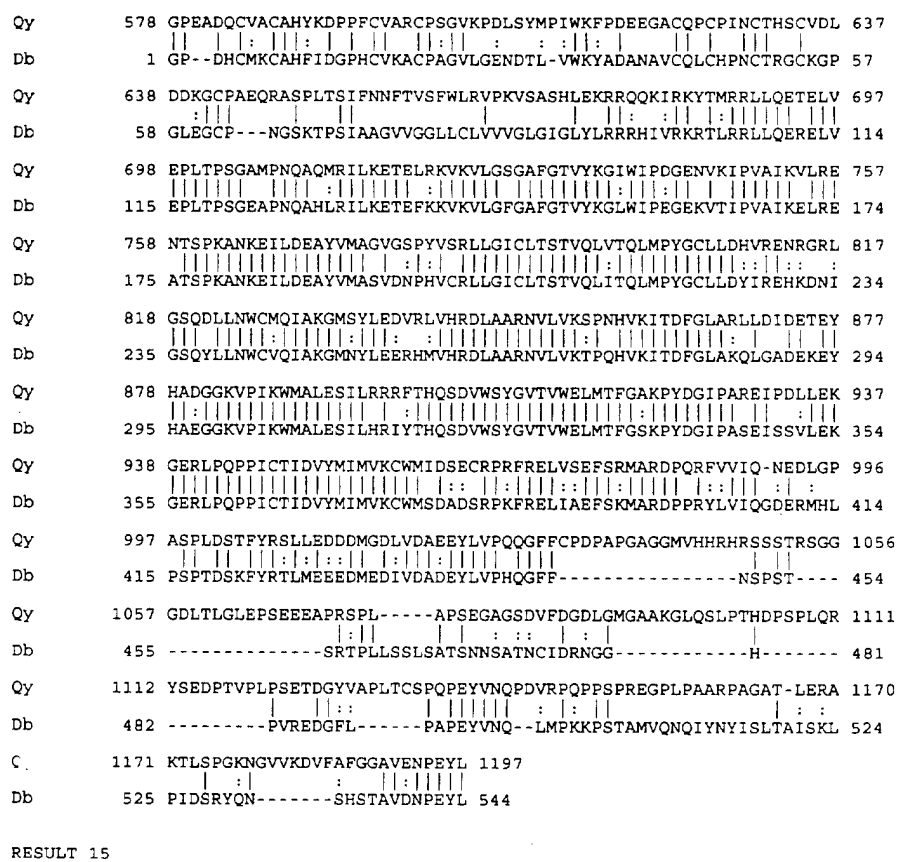
RESULT 12
TVYUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Mart
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specif
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

```
  Query Match             24.3%;  Score 1658;  DB 1;  Length 604;
  Best Local Similarity   51.4%;  Pred. No. 2e-61;
  Matches  354;  Conservative   69;  Mismatches  142;  Indels  124;  Gaps   14;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
              |||: | | ||  ||:||  :   : : |: |    || |  ||| |      :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAM 706
                |   ||          |       |  ||:  :|| |:|||||| ||||||||||
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEA 118

Qy        707 PNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 766
              |||| :|||||||| :|||||||||||:|||:|||:|| |||||||| ||| ||||||||
Db        119 PNQAHLRILKETEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKELREATSPKANKE 178

Qy        767 ILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWC 826
              |||||||| | :|:| |||||||||||||:|||||||||::||::  :||| |||||
Db        179 ILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWC 238

Qy        827 MQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPI 886
              :||||||:|||: |||||||||||||||:| |||||||||:||  || ||||:||||||
Db        239 VQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPI 298

Qy        887 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 946
              |||||||| | :|||||||||||||||||||:|||||||| ||  :|||||||||||||
Db        299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy        947 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1005
              |||||||||||||||:: ||:||||::|||:|||||  |::||| :| :   || || ||
Db        359 CTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy       1006 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1065
              |:|:|::|| |:|||:|||| ||||                | ||
Db        419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------- 449

Qy       1066 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1120
                 |:||      | |  : ::  |         :  |  |  : | :|||| |||
Db        450 -----SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNF 496

Qy       1121 PSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKN 1178
                |:  ||::      | ||||||  : |: ||
Db        497 LEESIDDGFL------PAPEYVNQ--LMPKKPSTAM----------------------- 524

Qy       1179 GVVKDVFAF-------------------GGAVENPEYLTPQGGAAPQPHPPPAFSPAFDN 1219
                |   :: |                 ||:|||||          :  |      |::
Db        525 -VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVPES 575

Qy       1220 LYYWDQDPPERGAPPSTFKGTPTAENPEY 1248
                || |    :              :||:|
Db        576 SPYWIQSGNHQ----------INLDNPDY 594


RESULT 13
GQFFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <LIV>
A;Cross-references: EMBL:K03054
R;Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021
A;Molecule type: DNA
A;Residues: 'A',832-866,'V',868-943,'QTPSLVK' <WAD>
A;Cross-references: EMBL:X02293; NID:g7922; PIDN:CAA26157.1; PID:g929565
C;Comment: This sequence is tentative because the introns have not been identified.
C;Genetics:
A;Gene: FlyBase:Egfr
A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2 57F
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TMM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

  Query Match             23.6%;  Score 1611.5;  DB 1;  Length 1330;
  Best Local Similarity   29.6%;  Pred. No. 3.5e-59;
  Matches  411;  Conservative  175;  Mismatches  414;  Indels  387;  Gaps   40;

Qy         80 VQGYVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASP 134
              :  |::|  : :      ||:|:||  ||     |: ||| |                :
Db         38 ITNYIVIGLDLIPCTLSYRLQIIRGRTLFSLSVEEEKYALFV----------------TY 81

Qy        135 GGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHP 194
                :  |::  | ::| | |    |  ||:  || | :|                |  | |
Db         82 SKMYTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPK 141

Qy        195 CSPMCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDC 252
              |   |    ||||  ::||  ::  |:  ||  || || | :|||  || |||||   ||
Db        142 CHESCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDC 200

Qy        253 LACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCT 312
              :|| :|    : :  || :  ||  |:    ||||:| :||:||  ||  :|  | |:|
Db        201 IACKNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACV 259

Qy        313 LVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAG 366
                ||    :  |     |  |: || : |              |:| |       |
Db        260 RSCPQDKMDKGGE-----CVPCNGPCPKTC-------------PGVTVLHAGNIDSFRN 300

Qy        367 CKKIFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDS 418
              |  | |::  | ::| |  |  :|        || ||: :|| |::|||||| |
Db        301 CTVIDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQ 360

Qy        419 LPDLSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFV 477
              :|| |:||: | || |    : :| :    :  | :|:|::: ||  :| ||  ||:|
Db        361 FRNLSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYV 420

Qy        478 HTVPWDQLFRNPHQALLHTANRPEDEC-------------------------------- 504
              : |  : : | |  :     | | |
Db        421 SNIRWPAIQKEPEQKVWVNENLRADLCGKFLTILISVQHNIIMHIFAICREKWNHLLGSV 480

Qy        505 ------------------------------------------------------------ 504

Db        481 QRGRLLGSWHGSVPYLQELQFQWHLHRRLWLYIQVSINSTQDKSNEHQLTDACYSPSVPT 540

Qy        505 ------------------------------------------------------------VG 506
                                                                           :
Db        541 SLTIERARYAIQSAGLAMELEQITARSASMRHSKTLPAEGRQVPRWVFLGVCASARAGIA 600

Qy        507 EGLA-------CHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYV---N 556
              | ||       || ||    :|     |  |: : | ::|  ||      | ::
Db        601 EPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCETEC------PADHYTDEE 654

Qy        557 ARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYK------DPPF-------CVARCPSG 603
              | |   ||||   ||  | || || | :| ::|       |:       | ::||
Db        655 QRECFQRHPEC---NG---CTGPGADDCKSCRNFKLFDANETGPYVNSTMFNCTSKCPLE 708

Qy        604 VK-PDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSF 662
```



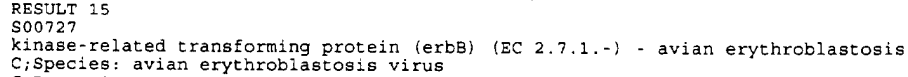
RESULT 14
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
R;Vennstroem, B.
submitted to the EMBL Data Library, March 1993
A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X12707
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 23.4%; Score 1602; DB 2; Length 544;
Best Local Similarity 54.1%; Pred. No. 3.6e-59;
Matches 339; Conservative 63; Mismatches 135; Indels 90; Gaps 13;



RESULT 15
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mu
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

Query Match 23.3%; Score 1595; DB 2; Length 545;
Best Local Similarity 54.1%; Pred. No. 7e-59;
Matches 339; Conservative 62; Mismatches 136; Indels 90; Gaps 13;



```
Qy        638 DDKGCPAEQRASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELV 697
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELV 114

Qy        698 EPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRE 757
Db        115 EPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELRE 174

Qy        758 NTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRL 817
Db        175 ATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNI 234

Qy        818 GSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEY 877
Db        235 GSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQLGADEKEY 294

Qy        878 HADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 937
Db        295 HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354

Qy        938 GERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGP 996
Db        355 GERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHL 414

Qy        997 ASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGG 1056
Db        415 PSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST---- 454

Qy       1057 GDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQR 1111
Db        455 --------------SRTPLLSSLSATSNNSATNCIDRNGG-----------H------- 481

Qy       1112 YSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERA 1170
Db        482 ---------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAISKL 524

Qy       1171 KTLSPGKNGVVKDVFAFGGAVENPEYL 1197
Db        525 PMDSRYQN-------SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:27:28
Job time : 30.2855 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 Seconds
(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-632-652-14
Perfect score: 6834
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6623	96.9	1255	1	ERB2_HUMAN	P04626 homo sapien
2	5865	85.8	1257	1	ERB2_RAT	P06494 rattus norv
3	5841.5	85.5	1254	1	ERB2_MESAU	Q60553 mesocricetu
4	3111	45.5	1210	1	EGFR_HUMAN	P00533 homo sapien
5	3081	45.1	1210	1	EGFR_MOUSE	Q01279 mus musculu
6	2920.5	42.7	1308	1	ERB4_HUMAN	Q15303 homo sapien
7	2906	42.5	1308	1	ERB4_RAT	Q62956 rattus norv
8	2643.5	38.7	1167	1	XMRK_XIPMA	P13388 xiphophorus
9	2398.5	35.1	1342	1	ERB3_HUMAN	P21860 homo sapien
10	2325.5	34.0	1339	1	ERB3_RAT	Q62799 rattus norv
11	1936	28.3	1426	1	EGFR_DROME	P04412 drosophila
12	1704.5	24.9	634	1	ERBB_ALV	P00534 avian leuko
13	1658	24.3	604	1	ERBB_AVIER	P00535 avian eryth
14	1585	23.2	540	1	ERBB_AVIEU	P11273 avian eryth
15	1550	22.7	703	1	EGFR_CHICK	P13387 gallus gall
16	1274	18.6	1323	1	LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.7	245	1	ERB2_MOUSE	P70424 mus musculu
18	713	10.4	1363	1	ILPR_BRALA	O02466 branchiosto
19	693.5	10.1	1372	1	INSR_MOUSE	P15208 mus musculu
20	692	10.1	1300	1	IRR_MOUSE	Q9wtl4 mus musculu
21	691	10.1	1382	1	INSR_HUMAN	P06213 homo sapien
22	690	10.1	1383	1	INSR_RAT	P15127 rattus norv
23	687	10.1	1297	1	IRR_HUMAN	P14616 homo sapien
24	680.5	10.0	1300	1	IRR_CAVPO	P14617 cavia porce
25	680	10.0	1477	1	HTK7_HYDAT	Q25197 hydra atten
26	680	10.0	1607	1	MIPR_LYMST	Q25410 lymnaea sta
27	644	9.4	1367	1	IG1R_HUMAN	P08069 homo sapien
28	630	9.2	1373	1	IG1R_MOUSE	Q60751 mus musculu
29	626.5	9.2	1370	1	IG1R_RAT	P24062 rattus norv
30	617	9.0	1390	1	INSR_AEDAE	Q93105 aedes aegyp
31	616	9.0	2146	1	INSR_DROME	P09208 drosophila
32	591	8.6	1114	1	RET_HUMAN	P07949 homo sapien
33	589.5	8.6	1053	1	FAK1_CHICK	Q00944 gallus gall
34	582	8.5	757	1	HT16_HYDAT	P53356 hydra atten
35	579	8.5	1052	1	FAK1_MOUSE	P34152 mus musculu
36	578.5	8.5	984	1	EPB1_CHICK	Q07494 gallus gall
37	576.5	8.4	984	1	EPB1_RAT	P09759 rattus norv
38	576	8.4	1055	1	FAK1_RAT	O35346 rattus norv
39	572	8.4	1052	1	FAK1_HUMAN	Q05397 homo sapien
40	570.5	8.3	984	1	EPB1_HUMAN	P54762 homo sapien
41	570	8.3	987	1	EPB4_HUMAN	P54760 homo sapien
42	569	8.3	902	1	EPBB_XENLA	Q91736 xenopus lae
43	569	8.3	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	555.5	8.1	1009	1	FAK2_RAT	P70600 rattus norv
45	554.5	8.1	746	1	ABL_MLVAB	P00521 abelson mur

ALIGNMENTS

RESULT 1
ERB2_HUMAN
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

```
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
CC       POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC       OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC       ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M11767; AAA35808.1; -.
DR   EMBL; M11761; AAA35808.1; JOINED.
DR   EMBL; M11762; AAA35808.1; JOINED.
DR   EMBL; M11763; AAA35808.1; JOINED.
DR   EMBL; M11764; AAA35808.1; JOINED.
DR   EMBL; M11765; AAA35808.1; JOINED.
DR   EMBL; M11766; AAA35808.1; JOINED.
DR   EMBL; M11730; AAA75493.1; -.
DR   EMBL; M12036; AAA35978.1; -.
DR   EMBL; X03363; CAA27060.1; -.
DR   PIR; A25491; A25491.
DR   PIR; A24571; A24571.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3430; ERBB2.
DR   MIM; 164870; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Polymorphism.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1255       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    220    227       BY SIMILARITY.
FT   DISULFID    224    235       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     654    654       I -> V.
FT                                 /FTId=VAR_004077.
FT   VARIANT     655    655       I -> V.
FT                                 /FTId=VAR_004078.
FT   CONFLICT   1170   1170       P -> A (IN REF. 2).
SQ   SEQUENCE   1255 AA;  137909 MW;  39E9DFDA04DCF962 CRC64;

  Query Match             96.9%;  Score 6623;  DB 1;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              : |  |    | ||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
```

```
Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 2
ERB2_RAT
ID   ERB2_RAT       STANDARD;      PRT;  1257 AA.
AC   P06494;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE   receptor-related protein).
GN   ERBB2 OR NEU.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=86118662; PubMed=3945311;
RA   Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT   "The neu oncogene encodes an epidermal growth factor receptor-related
RT   protein.";
RL   Nature 319:226-230(1986).
RN   [2]
RP   SEQUENCE OF 852-905 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   STRUCTURE BY NMR OF 650-668.
RX   MEDLINE=92155181; PubMed=1346763;
RA   Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,
RA   Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT   "Three dimensional structure of the transmembrane region of the proto-
RT   oncogenic and oncogenic forms of the neu protein.";
RL   EMBO J. 11:43-48(1992).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC       THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03362; CAA27059.1; ALT_INIT.
DR   PIR; A24562; TVRTNU.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1257       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    677       POTENTIAL.
FT   DOMAIN      678   1257       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      159    369       CYS-RICH.
FT   DOMAIN      473    646       CYS-RICH.
FT   DOMAIN      722    989       PROTEIN KINASE.
FT   NP_BIND     728    736       ATP (BY SIMILARITY).
FT   BINDING     755    755       ATP (BY SIMILARITY).
FT   ACT_SITE    847    847       BY SIMILARITY.
FT   DISULFID    196    205       BY SIMILARITY.
FT   DISULFID    200    213       BY SIMILARITY.
FT   DISULFID    221    228       BY SIMILARITY.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    237    245       BY SIMILARITY.
FT   DISULFID    241    253       BY SIMILARITY.
FT   DISULFID    256    265       BY SIMILARITY.
FT   DISULFID    269    296       BY SIMILARITY.
FT   DISULFID    300    312       BY SIMILARITY.
FT   DISULFID    316    332       BY SIMILARITY.
FT   DISULFID    335    339       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    578       BY SIMILARITY.
```

```
FT   DISULFID    569    586       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    602    625       BY SIMILARITY.
FT   DISULFID    628    636       BY SIMILARITY.
FT   DISULFID    632    644       BY SIMILARITY.
FT   MOD_RES    1141   1141       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1250   1250       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    260    260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     661    661       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1257 AA;  138831 MW;  6129264583011402 CRC64;

  Query Match             85.8%;  Score 5865;  DB 1;  Length 1257;
  Best Local Similarity   86.2%;  Pred. No. 1.6e-302;
  Matches 1084;  Conservative   48;  Mismatches  123;  Indels    2;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| |||||||||||||||:|||||||::|||||||||||||||| ||||||||
Db         61 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||||: |||||||| :||||:| |||||
Db        121 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 180

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |   |||||||| ||:| || : |||| |||| || |:|  ||||||| |||||||||
Db        181 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 240

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||| |
Db        241 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 300

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||||||:::::
Db        301 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 360

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
              || |||||||||||||||||||:|  |||:|||||||||||||||||||||||||||||
Db        361 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 420

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              ||||||||::|||||||:|||||||||||  |||||||||||||||| | |||||||
Db        421 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 480

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||||: |||||:: ||  || |: ||| |||||||||||||| |||||
Db        481 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : |||||||||||| | |||| |||| |||||||   |||
Db        541 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 600

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNF 658
              |||||||||||||||:||||| |||||||||||||::|||||||||:| |
Db        601 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRASPVTFIIATV 660

Qy        659 TVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 718
                    :  |    | |||:|||||||||||||||||||||||||||||||||||||
Db        661 VGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 720

Qy        719 ELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVG 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVG 780

Qy        779 SPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLED 838
              |||||||||||||||||||||||||||||||||||||||||||:|||||||||||:|||||
Db        781 SPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYLED 840

Qy        839 VRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRR 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 VRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRR 900

Qy        899 FTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKC 958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 FTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKC 960

Qy        959 WMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLV 1018
              ||||||||||||||||||||||||||||||||||||:||:||||||||||||||||||||
Db        961 WMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDDDMGDLV 1020

Qy       1019 DAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPS 1078
              |||||||||||||| ||| || |  | ||||||||||:|||||||||  ||||||||
Db       1021 DAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPPRSPLAPS 1080

Qy       1079 EGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPE 1138
              |||||||||||| ||  ||||||  || ||||||||||:||| |||||||| ||||||
Db       1081 EGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPLACSPQPE 1140

Qy       1139 YVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLT 1198
              |||| :|:|||  |||||  |||||||| ||||||||||||||||||||||||||||||
Db       1141 YVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLV 1200

Qy       1199 PQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |: | |  ||| |||||||||||||:  |:| ||| |:|||||||||||||||||
Db       1201 PREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGLDVPV 1257


RESULT 3
ERB2_MESAU
ID   ERB2_MESAU     STANDARD;      PRT;  1254 AA.
AC   Q60553;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN   ERBB2 OR NEU.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Nerve;
RX   MEDLINE=94193007; PubMed=7908275;
RA   Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA   Yamazaki Y., Ishikawa T.;
RT   "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL   Gene 140:251-255(1994).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D16295; BAA03801.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1254       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1254       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      158    368       CYS-RICH.
FT   DOMAIN      472    644       CYS-RICH.
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1247   1247       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     658    658       V -> E (IN ONCOGENIC NEU).
FT   VARIANT     659    659       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1254 AA;  138252 MW;  974C3791C21F2BE1 CRC64;

  Query Match             85.5%;  Score 5841.5;  DB 1;  Length 1254;
  Best Local Similarity   85.7%;  Pred. No. 2.8e-301;
  Matches 1076;  Conservative   57;  Mismatches  121;  Indels    1;  Gaps    1;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||  | |||||||| |||: |||||||||||||||||||||::|||||||||||||
Db          1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:||||:||| |||||||||||||||| ||||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||:| |  || :| |||||||||||||||||||: |||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                |||||||| ||:| || : ||| | ||||:|| |:        :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||:|||||||||:|||||||||||||||||| ||||||||||::::
Db        301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||:|:|  ||| |||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||:||||||:||:||||| |||||| |||||||||||||| ||| ||||||||||
Db        421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||||: |  |:||  :  ||: ||| |||||||||||||| ||||||||
Db        481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |:|||| :|||||||| :|||||||||||| : || | ||||| || |||| |||||||
Db        541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||:|||| |||||||||||||||::|||||||||| |||
Db        601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                  |  |    | |||:||||||||||||||||||||||||||||||||||||||||
Db        661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGLGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              |||||||||||||||||||||||||||||||:|||||||||||||:|||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              |||||||||| ||||||||   | ||||||||||:||||:||| || ||||||||||
Db       1021 EEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||:|:||||| || ||:   | ||||||||||:|||:||||||||| ||||||||
Db       1081 AGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLACSPQPEYV 1140
```

```
Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
             |||:||||||   |||||  |||||||| |||||||||||||| ||||||||||| |:
Db       1141 NQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVENPEYLVPR 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              ||:| ||| |||  ||||||||||||| |||:||:||:||||||||||||||||
Db       1201 GGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDVPV 1254


RESULT 4
EGFR_HUMAN
ID   EGFR_HUMAN     STANDARD;      PRT;  1210 AA.
AC   P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC   O00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE   protein-tyrosine kinase ErbB-1).
GN   EGFR OR ERBB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=84219729; PubMed=6328312;
RA   Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA   Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA   Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT   "Human epidermal growth factor receptor cDNA sequence and aberrant
RT   expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL   Nature 309:418-425(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=95382957; PubMed=7654368;
RA   Ilekis J.V., Stark B.C., Scoccia B.;
RT   "Possible role of variant RNA transcripts in the regulation of
RT   epidermal growth factor receptor expression in human placenta.";
RL   Mol. Reprod. Dev. 41:149-156(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97078686; PubMed=8918811;
RA   Reiter J.L., Maihle N.J.;
RT   "A 1.8 kb alternative transcript from the human epidermal growth
RT   factor receptor gene encodes a truncated form of the receptor.";
RL   Nucleic Acids Res. 24:4050-4056(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97256547; PubMed=9103388;
RA   Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT   "Expression of a truncated epidermal growth factor receptor-like
RT   protein (TEGFR) in ovarian cancer.";
RL   Gynecol. Oncol. 65:36-41(1997).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC   TISSUE=Placenta;
RX   MEDLINE=21100872; PubMed=11161793;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative EGFR transcripts encoding truncated receptor
RT   isoforms.";
RL   Genomics 71:1-20(2001).
RN   [6]
RP   SEQUENCE OF 575-687 FROM N.A.
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Human and mouse alternative EGFR transcripts encoding only the
RT   extracellular domain of the receptor.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 713-924 FROM N.A.
RX   MEDLINE=84196372; PubMed=6326261;
RA   Lin C.R., Chen W.S., Kruiger W., Stolarsky L.S., Weber W.,
RA   Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT   "Expression cloning of human EGF receptor complementary DNA: gene
RT   amplification and three related messenger RNA products in A431
RT   cells.";
RL   Science 224:843-848(1984).
RN   [8]
RP   SEQUENCE OF 150-962 FROM N.A.
RX   MEDLINE=84245835; PubMed=6330563;
RA   Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA   Roe B.A., Merlino G.T., Pastan I.;
RT   "Human epidermal growth factor receptor cDNA is homologous to a
RT   variety of RNAs overproduced in A431 carcinoma cells.";
RL   Nature 309:806-810(1984).
RN   [9]
RP   SEQUENCE OF 1028-1210 FROM N.A.
RX   MEDLINE=85046483; PubMed=6093780;
RA   Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA   O'Malley B.W.;
RT   "Isolation of an evolutionarily conserved epidermal growth factor
RT   receptor cDNA from human A431 carcinoma cells.";
RL   Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN   [10]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=88217333; PubMed=3329716;
RA   Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA   Waterfield M.D.;
RT   "The human EGF receptor gene: structure of the 110 kb locus and
RT   identification of sequences regulating its transcription.";
RL   Oncogene Res. 1:375-396(1987).
RN   [11]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=91107677; PubMed=1988448;
RA   Haley J.D., Waterfield M.D.;
RT   "Contributory effects of de novo transcription and premature
RT   transcript termination in the regulation of human epidermal growth
RT   factor receptor proto-oncogene RNA synthesis.";
RL   J. Biol. Chem. 266:1746-1753(1991).
RN   [12]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=85270438; PubMed=2991899;
RA   Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT   "Characterization and sequence of the promoter region of the human
RT   epidermal growth factor receptor gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN   [13]
RP   SEQUENCE OF 540.
RA   Kohda D.;
RL   Submitted (SEP-1997) to the SWISS-PROT data bank.
RN   [14]
RP   RECEPTOR ACTIVITY.
RX   MEDLINE=84191554; PubMed=6325948;
RA   Mroczkowski B., Mosig G., Cohen S.;
RT   "ATP-stimulated interaction between epidermal growth factor receptor
RT   and supercoiled DNA.";
RL   Nature 309:270-273(1984).
RN   [15]
RP   PHOSPHORYLATION.
RX   MEDLINE=89278137; PubMed=2543678;
RA   Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA   Howk R., Givol D., Ullrich A., Schlessinger J.;
RT   "All autophosphorylation sites of epidermal growth factor (EGF)
RT   receptor and HER2/neu are located in their carboxyl-terminal tails.
RT   Identification of a novel site in EGF receptor.";
RL   J. Biol. Chem. 264:10667-10671(1989).
RN   [16]
```

```
RP   CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
RP   ASN-528.
RX   MEDLINE=96398132; PubMed=8962717;
RA   Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
RT   "Analysis of the glycosylation patterns of the extracellular domain of
RT   the epidermal growth factor receptor expressed in Chinese hamster
RT   ovary fibroblasts.";
RL   Growth Factors 13:121-132(1996).
RN   [17]
RP   CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
RP   ASN-603.
RX   MEDLINE=20198209; PubMed=10731668;
RA   Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
RT   "Characterization of the N-oligosaccharides attached to the atypical
RT   Asn-X-Cys sequence of recombinant human epidermal growth factor
RT   receptor.";
RL   J. Biochem. 127:65-72(2000).
RN   [18]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=98225196; PubMed=9556602;
RA   Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
RT   "Disulfide bond structure of human epidermal growth factor receptor.";
RL   J. Biol. Chem. 273:11150-11157(1998).
RN   [19]
RP   REVIEW.
RX   MEDLINE=87297456; PubMed=3039909;
RA   Carpenter G.;
RT   "Receptors for epidermal growth factor and other polypeptide
RT   mitogens.";
RL   Annu. Rev. Biochem. 56:881-914(1987).
CC   -!- FUNCTION: Receptor for EGF, but also for other members of the EGF
CC       family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
CC       EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
CC       involved in the control of cell growth and differentiation.
CC   -!- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
CC       secreted.
CC   -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
CC       truncated isoform/TEGFR, 3/p110 and 4; are produced by
CC       alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
CC       expressed in ovarian cancers.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X00588; CAA25240.1; -.
DR   EMBL; U95089; AAB53063.1; -.
DR   EMBL; U48722; AAC50802.1; -.
DR   EMBL; U48723; AAC50804.1; -.
DR   EMBL; U48724; AAC50796.1; -.
DR   EMBL; U48725; AAC50797.1; -.
DR   EMBL; U48726; AAC50798.1; -.

  Query Match             45.5%;  Score 3111;  DB 1;  Length 1210;
  Best Local Similarity   49.5%;  Pred. No. 4.3e-157;
  Matches  625;  Conservative  172;  Mismatches  364;  Indels  102;  Gaps   21;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
              || ||  |    :|| ||  ||    : | |   |: ::  |:|| ||||:||:  |
Db         14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
              ||||: |||| |||||| | | ::||: |:|:||   :|::|||||| | |
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
                |: .||:|| :|:| ||| | |    || ||  ::| |:||   :    :::   |
Db        127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy        189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
              :|  | | |    |||   |:|| ||: :||  |: ||:|  |:|||| |||||||||
Db        184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy        248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
              : |||| |  |       |:  || |: ||  |::  ||||:|:|||:|| || ||: ||
Db        244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy        308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
              |||   |   : |:  ||| ::|:||  || :|| |:|:   | :|  |  | :: |
Db        304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy        367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
              |  | | |  || :| ||  ::| || |::| : :|::||||:| | |||::  ||  |:
Db        362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy        427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
              ||::|||   :| :|| :  | |: |||||:|:  |  :|  | :||: :|: | :||
Db        422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy        487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
                    :|| |: |   |  || ||:   |||| |  ||:|   ||:|||::| :
Db        482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNL 541

Qy        547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606
              |:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||  ||:||
Db        542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy        607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSFWLRV 666
              : : : :||: |    |  | |||: |     :|||       | ||       : | :
Db        602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLLL 658

Qy        667 PKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVL 726
                  |  ||:  :|| |:|||||| |||||||||  |||| :|||||| :|:|||
Db        659 VVALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVL 718

Qy        727 GSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLL 786
              |||||||||||:|||:|| |||||||| ||| |||||||||||||||| | :|:| |||
Db        719 GSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLL 778

Qy        787 GICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDL 846
              ||||||||||:|||||:|||||:|||::  :||| |||||:||||||:|||| |||||||
Db        779 GICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDL 838

Qy        847 AARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVW 906
              ||||||||:| |||||||||:||  :| |||||:|||||||||||||| | :|||||||
Db        839 AARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVW 898

Qy        907 SYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECR 966
              |||||||||||||:|||||||  ||  :|||||||||||||||||||||||||:: |
Db        899 SYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSR 958

Qy        967 PRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLV 1025
              |:|||: |||:|||||||::||| :| :   || || |||:|::::|| |:|||:|||:
Db        959 PKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLI 1018

Qy       1026 PQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDV 1085
              |||||                                  |     |:||  |  | |:
Db       1019 PQQGFF----------------------------------SSPSTSRTPLLSSLSATSN- 1043

Qy       1086 FDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQP 1143
                :  :      |||| |  : | ||||| |||  |  ::  | ::      | |||:||
Db       1044 -NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------PVPEYINQ- 1095
```

```
Qy       1144 DVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-TPQGG 1202
                |  :|    : |:   :|          :| ::   :|      || ||||| | |
Db       1096 SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNPEYLNTVQ-- 1143

Qy       1203 AAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTPTAENP 1246
                     |   :  ||:  :| |         :|       |:   |   |||: ||||
Db       1144 -------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGS-TAENA 1195

Qy       1247 EYL 1249
                 |||
Db       1196 EYL 1198
```

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC ---

CC ---
DR EMBL; X78987; CAA55587.1; -.
DR EMBL; U03425; AAA17899.1; -.
DR EMBL; X59698; CAA42219.1; -.
DR EMBL; L06864; AAA53029.1; -.
DR EMBL; Z12608; CAA78249.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.

FT	SIGNAL	1	24	POTENTIAL.
FT	CHAIN	25	1210	EPIDERMAL GROWTH FACTOR RECEPTOR.
FT	DOMAIN	25	647	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	648	670	POTENTIAL.
FT	DOMAIN	671	1210	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	75	300	APPROXIMATE.
FT	REPEAT	390	600	APPROXIMATE.
FT	DOMAIN	1028	1071	SER-RICH.
FT	DOMAIN	714	981	PROTEIN KINASE.
FT	NP_BIND	720	728	ATP (BY SIMILARITY).
FT	BINDING	747	747	ATP (BY SIMILARITY).
FT	ACT_SITE	839	839	BY SIMILARITY.
FT	DISULFID	190	199	BY SIMILARITY.
FT	DISULFID	194	207	BY SIMILARITY.
FT	DISULFID	215	223	BY SIMILARITY.
FT	DISULFID	219	231	BY SIMILARITY.
FT	DISULFID	232	240	BY SIMILARITY.
FT	DISULFID	236	248	BY SIMILARITY.
FT	DISULFID	251	260	BY SIMILARITY.
FT	DISULFID	264	291	BY SIMILARITY.
FT	DISULFID	295	307	BY SIMILARITY.
FT	DISULFID	311	326	BY SIMILARITY.
FT	DISULFID	329	333	BY SIMILARITY.
FT	DISULFID	506	515	BY SIMILARITY.
FT	DISULFID	510	523	BY SIMILARITY.
FT	DISULFID	526	535	BY SIMILARITY.
FT	DISULFID	539	555	BY SIMILARITY.
FT	DISULFID	558	571	BY SIMILARITY.
FT	DISULFID	562	579	BY SIMILARITY.
FT	DISULFID	582	591	BY SIMILARITY.
FT	DISULFID	595	617	BY SIMILARITY.
FT	DISULFID	620	628	BY SIMILARITY.
FT	DISULFID	624	636	BY SIMILARITY.
FT	MOD_RES	680	680	PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT	MOD_RES	1092	1092	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```
FT   MOD_RES     1110   1110       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     1172   1172       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     1197   1197       PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT                                 (BY SIMILARITY).
FT   CARBOHYD     128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     175    175       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     352    352       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     444    444       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      19     19       C -> S (IN REF. 2).
FT   CONFLICT     539    539       C -> W (IN REF. 5).
FT   CONFLICT     991    991       L -> F (IN REF. 4).
FT   CONFLICT    1116   1117       HP -> DR (IN REF. 6).
SQ   SEQUENCE   1210 AA;  134853 MW;  690E20D46DF2D2F5 CRC64;

  Query Match           45.1%;  Score 3081;  DB 1;  Length 1210;
  Best Local Similarity  49.3%;  Pred. No. 1.6e-155;
  Matches  627;  Conservative  164;  Mismatches  371;  Indels  110;  Gaps   23;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSF 662
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 WLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRK 722
Db        657 LFIVVVALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKK 716

Qy        723 VKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYV 782
Db        717 IKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHV 776

Qy        783 SRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLV 842
Db        777 CRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLV 836

Qy        843 HRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQ 902
Db        837 HRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQ 896

Qy        903 SDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMID 962
Db        897 SDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVKCWMID 956

Qy        963 SECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAE 1021
Db        957 ADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVVDAD 1016

Qy       1022 EYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGA 1081
Db       1017 EYLIPQQGFF----------------NSPST------------------SRTPLLSSLSA 1042

Qy       1082 GSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEY 1139
Db       1043 TSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------PVPEY 1092

Qy       1140 VNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-T 1198
Db       1093 VNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNPEYLNT 1141

Qy       1199 PQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTPT 1242
Db       1142 AQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIFKG-PT 1191

Qy       1243 AENPEYLGLDVP 1254
Db       1192 AENAEYLRVAPP 1203
```

```
RESULT 6
ERB4_HUMAN
ID   ERB4_HUMAN     STANDARD;      PRT;  1308 AA.
AC   Q15303;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE   (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN   ERBB4 OR HER4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM JM-A).
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=93189574; PubMed=8383326;
RA   Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA   Foy L., Neubauer M.G., Shoyab M.;
RT   "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT   epidermal growth factor receptor family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97476287; PubMed=9334263;
RA   Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA   Klagsbrun M.;
RT   "A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
RT   tissue distribution and differential processing in response to
RT   phorbol ester.";
RL   J. Biol. Chem. 272:26761-26768(1997).
```

```
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC       FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC       BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC       NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC       CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC       KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC       PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC       LUNG, SALIVARY GLAND, AND PANCREAS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07868; AAB59446.1; -.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3432; ERBB4.
DR   MIM; 600543; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN      676   1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      186    334       CYS-RICH.
FT   DOMAIN      496    633       CYS-RICH.
FT   DOMAIN      718    985       PROTEIN KINASE.
FT   NP_BIND     724    732       ATP (BY SIMILARITY).
FT   BINDING     751    751       ATP (BY SIMILARITY).
FT   ACT_SITE    843    843       BY SIMILARITY.
FT   DISULFID    189    197       BY SIMILARITY.
FT   DISULFID    193    205       BY SIMILARITY.
FT   DISULFID    213    221       BY SIMILARITY.
FT   DISULFID    217    229       BY SIMILARITY.
FT   DISULFID    230    238       BY SIMILARITY.
FT   DISULFID    234    246       BY SIMILARITY.
FT   DISULFID    249    258       BY SIMILARITY.
FT   DISULFID    262    289       BY SIMILARITY.
FT   DISULFID.   293    304       BY SIMILARITY.
FT   DISULFID    308    323       BY SIMILARITY.
FT   DISULFID    326    330       BY SIMILARITY.
FT   DISULFID    503    512       BY SIMILARITY.
FT   DISULFID    507    520       BY SIMILARITY.
FT   DISULFID    523    532       BY SIMILARITY.
FT   DISULFID    536    552       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    559    577       BY SIMILARITY.
FT   DISULFID    580    589       BY SIMILARITY.
FT   DISULFID    593    614       BY SIMILARITY.
FT   DISULFID    617    625       BY SIMILARITY.
FT   DISULFID    621    633       BY SIMILARITY.
FT   MOD_RES    1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    626    648       NGPTSHDCIYYPWTGHSTLPQHA -> IGSSIEDCIGLMD
FT                                 (IN ISOFORM JM-B).
SQ   SEQUENCE   1308 AA;  146807 MW;  5E4AE80985D88761 CRC64;

  Query Match           42.7%;  Score 2920.5;  DB 1;  Length 1308;
  Best Local Similarity 44.6%;  Pred. No. 5.4e-147;
  Matches  604;  Conservative  178;  Mismatches  384;  Indels  187;  Gaps   26;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  |  ||||: ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::||||| ||| : :|  ||| ||| |:|| |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||     ||  ||:   |   :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||     ||  |:| : |: |:  | : | |:|   : :          :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :    : ||:| || |: ||||:| | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||   :| ::   ||  : |  ||: |: ||:   |||||| ||::| :| ||: |:|
Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              | :  |  ||: |   |: | |:|:  ::| :|| ||  | |  |:|:|| | || :||
```

```
Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIF----NNF 658
              |::   |:  |:|: | :  | ||  ||| |       |           |:     :
Db        596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDC----------IYYPWTGHS 642

Qy        659 TVSFWLRVPKVSAS---------------HLEKRRQQKIRKYTMRRLLQETELVEPLTPS 703
              |:    | | ::|                 :  ||:  :|  :|| | ||||||||||
Db        643 TLPQHARTPLIAAGVIGGLFILVIVGLTFAVYVRRKSIKKKRALRRFL-ETELVEPLTPS 701

Qy        704 GAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKA 763
              |  |||||:||||||||::||||||||||||||||:|:|| ||||||||:| | | |||
Db        702 GTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKA 761

Qy        764 NKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLL 823
              | | :||| :|| :  |:: ||||:||: |:|||||||:||||::| |::  :||| ||
Db        762 NVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLL 821

Qy        824 NWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGK 883
              |||:|||||| ||| |||||||||||||||||||||||||||: || ||:|||||
Db        822 NWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGK 881

Qy        884 VPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQ 943
              :|||||||| |  |:|||||||||||:|||||| |||||| ||||||||||||||||
Db        882 MPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQ 941

Qy        944 PPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDS 1002
              |||||||||:|||||||:: ||:|:|| :|||||||||||::||| :| :   || ||
Db        942 PPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDS 1001

Qy       1003 TFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLG 1062
               |:::||:::|: |::|||||||| | |  | |      :   | |  | ||      :|
Db       1002 KFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIG 1049

Qy       1063 LEPSEEEAPRS---------------------------PLAP-SEGAGSDVFDGDLGMG 1093
                |     | |                          | || ::|| :::||     |
Db       1050 HSPPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNG 1109

Qy       1094 AAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVR 1146
                :   :    : |  |||| |||| |        : :||: |:   |: ||:|  :
Db       1110 TLRKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE-- 1167

Qy       1147 PQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQ 1206
                    | |  : |               ||| ::       |::||||     |
Db       1168 -------ENPFVSRR--------------KNGDLQ-------ALDNPEYHNASNG---- 1194

Qy       1207 PHPPPA---------------------------------FSPAFDNLYYWDQDPPERGA 1232
                | |                                  ||||  ||:   | |
Db       1195 --PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRST 1252

Qy       1233 --PPSTFKGTPT-------------AENPEYL 1249
                  |   :  |              |||||||
Db       1253 LQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285
```

```
RESULT 7
ERB4_RAT
ID   ERB4_RAT       STANDARD;      PRT;  1308 AA.
AC   Q62956; Q9Z2N7;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN   ERBB4 OR TYRO-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=98221155; PubMed=9553078;
RA   Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA   Marchionni M.A., Kelly R.A.;
RT   "Neuregulins promote survival and growth of cardiac myocytes.
RT   Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT   ventricular myocytes.";
RL   J. Biol. Chem. 273:10261-10269(1998).
RN   [2]
RP   SEQUENCE OF 848-901 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   SEQUENCE OF 1031-1198 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC       NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC       OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC       RETICULAR NUCLEUS OF THE THALLAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC       HEART.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF041838; AAD08899.1; -.
DR   EMBL; U52531; AAC53051.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN      676   1308       CYTOPLASMIC (POTENTIAL).
```

```
FT   DOMAIN      186    334       CYS-RICH.
FT   DOMAIN      496    633       CYS-RICH.
FT   DOMAIN      718    985       PROTEIN KINASE.
FT   NP_BIND     724    732       ATP (BY SIMILARITY).
FT   BINDING     751    751       ATP (BY SIMILARITY).
FT   ACT_SITE    843    843       BY SIMILARITY.
FT   DISULFID    189    197       BY SIMILARITY.
FT   DISULFID    193    205       BY SIMILARITY.
FT   DISULFID    213    221       BY SIMILARITY.
FT   DISULFID    217    229       BY SIMILARITY.
FT   DISULFID    230    238       BY SIMILARITY.
FT   DISULFID    234    246       BY SIMILARITY.
FT   DISULFID    249    258       BY SIMILARITY.
FT   DISULFID    262    289       BY SIMILARITY.
FT   DISULFID    293    304       BY SIMILARITY.
FT   DISULFID    308    323       BY SIMILARITY.
FT   DISULFID    326    330       BY SIMILARITY.
FT   DISULFID    503    512       BY SIMILARITY.
FT   DISULFID    507    520       BY SIMILARITY.
FT   DISULFID    523    532       BY SIMILARITY.
FT   DISULFID    536    552       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    559    577       BY SIMILARITY.
FT   DISULFID    580    589       BY SIMILARITY.
FT   DISULFID    593    614       BY SIMILARITY.
FT   DISULFID    617    625       BY SIMILARITY.
FT   DISULFID    621    633       BY SIMILARITY.
FT   MOD_RES    1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1062   1062       S -> N (IN REF. 3).
FT   CONFLICT   1080   1082       PYT -> SYR (IN REF. 3).
SQ   SEQUENCE   1308 AA;  146957 MW;  D944BB0996A08B41 CRC64;

  Query Match             42.5%;  Score 2906;  DB 1;  Length 1308;
  Best Local Similarity   44.5%;  Pred. No. 3.1e-146;
  Matches  603;  Conservative  185;  Mismatches  392;  Indels  174;  Gaps   26;

Qy          1 MELA-ALCRWGLLL--ALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQ 57
              |:||  |  || ||  |     :||  || ||: ||   :  |      ||  |: |:||
Db          1 MKLATGLWVWGSLLVAARTVQPSASQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVM 60

Qy         58 GNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVL 117
              ||||:|  :  |  ||||: |:|| ||||:|  || |  :||: |||:|||:|:|| ||||:
Db         61 GNLEITSIEHNRDLSFLRSIREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIF 120

Qy        118 DNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNN 177
              |     |          ||:|| |::||||| ||| : :|  ||| ||| |:||
Db        121 LNYRKDGNF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPW 171

Qy        178 QLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCC 236
              :||: |  |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||
Db        172 PSNMTLVSTIGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCC 230

Qy        237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
              | :|| ||:||| :|| ||:|| || |   ||    ||  ||:   |   :||:|| ||
Db        231 HRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCV 290

Qy        297 TACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF 356
              ||:|::  |  ||   ||    ||  |:| : |: |:  | : | |:|   : :
Db        291 KKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTV 348

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
                    :|  | || |:| ||      ||| :    : ||:| || |: ||||:| | ||
Db        349 DSSNIDKFINCTKINGNLIFLVTGIHGDPYNAIDAIDPEKLNVFRTVREITGFLNIQTWP 408

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
              ::  | ||| ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||:
Db        409 PNMTDFSVFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCY 468

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
              ||: |  ||    :| ::   ||  : |  ||: |: ||:   ||||| ||::| :| |
Db        469 YHTINWTTLFSTVNQRIVIRDNRRAENCTAEGMVCNHLCSNDGCWGPGPDQCLSCRRFSR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPF 595
              |: |:| | :  |  ||: |   |: |  :|:  ::| :|| ||  | |  |:|:|| |
Db        529 GKICIESCNLYDGEFREFENGSICVECDSQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPN 588

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIF 655
              || :||  ::   |:  |:|: |::  | ||  |||  |        |           |:
Db        589 CVEKCPDVLQGANSF--IFKYADQDRECHPCHPNCTQGCNGPTSHDC-----------IY 635

Qy        656 ----NNFTVSFWLRVPKVSAS--------------HLEKRRQQKIRKYTMRRLLQETEL 696
                  : |:    | | ::|                :  ||:   :|  :|| | ||||
Db        636 YPWTGHSTLPQHARTPLIAAGVIGGLFILVIMALTFAVYVRRKSIKKKRALRRFL-ETEL 694

Qy        697 VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLR 756
              |||||||  |||||:||||||||::||||||||||||||||:|:|| ||||||||:|
Db        695 VEPLTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILN 754

Qy        757 ENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGR 816
              | | |||  | :||| :|| |  |:: ||||:||: |:|||||||:||||::| |::
Db        755 ETTGPKANVEFMDEALIMASVDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDN 814

Qy        817 LGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETE 876
              :||| ||||:|||||| |||: ||||||||||||||||||||||||||||: || |
Db        815 IGSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKE 874

Qy        877 YHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLE 936
              |:|||||:|||||||| |  |:|||||||||||:|||||| ||||||| ||||||||
Db        875 YNADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLE 934

Qy        937 KGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LG 995
              ||||||||||||||::|||||||:: ||:|:|| :|||||||||::||| :| :
Db        935 KGERLPQPPICTIDVYIVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMK 994

Qy        996 PASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDP-------------------- 1035
              || || |:::||:::|: |::|||||||| | |  | |
Db        995 LPSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTSRTRIDSNRSEIGHSPP 1053

Qy       1036 ---APGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGM 1092
                  | :|    ::    :|: |   : :    :    | :|:|  :|| :::||
Db       1054 PAYTPMSGSQFVYQDGGFATQQG---MPMPYTATTSTIPEAPVA--QGATAEMFDDSCCN 1108

Qy       1093 GAAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDV 1145
              |  :        :  |||| |||| |         : :||: |:   |: ||:|  :
Db       1109 GTLRKPVVPHVQEDSSTQRYSADPTVFAPERNPRAELDEEGYMTPMHDKPKQEYLNPVE- 1167

Qy       1146 RPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAP 1205
                  | |  : |              ||| ::      |::|||| :   |
Db       1168 --------ENPFVSRR--------------KNGDLQ-------ALDNPEYHSASSG--- 1194

Qy       1206 QPHPPPA---------------------------------FSPAFDNLYYWDQDPPERG 1231
                 || |                                 ||||  ||:   | |
Db       1195 ---PPKAEDEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPDYWNHSLPPRS 1251

Qy       1232 A--PPSTFKGTPT--------------AENPEYL 1249
                 |   :  |               |||||||
Db       1252 TLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 8
XMRK_XIPMA
ID   XMRK_XIPMA     STANDARD;      PRT;  1167 AA.
AC   P13388;
```

```
DT   01-JAN-1990 (Rel. 13, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN   XMRK OR TU.
OS   Xiphophorus maculatus (Southern platyfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8083;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90015140; PubMed=2797166;
RA   Wittbrodt J., Adam D., Malitschek B., Maueler W., Raulf F.,
RA   Telling A., Robertson S.M., Schartl M.;
RT   "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT   inducing Tu locus in Xiphophorus.";
RL   Nature 341:415-421(1989).
RN   [2]
RP   REVISION TO 515.
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16891; CAA34770.2; -.
DR   PIR; S06142; S06142.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT   SIGNAL       1     25
FT   CHAIN       26   1167       MELANOMA RECEPTOR PROTEIN-TYROSINE
FT                                KINASE.
FT   DOMAIN       26    642       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    643    665       POTENTIAL.
FT   DOMAIN      666   1167       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      710    977       PROTEIN KINASE.
FT   NP_BIND     716    724       ATP (BY SIMILARITY).
FT   BINDING     743    743       ATP (BY SIMILARITY).
FT   ACT_SITE    835    835       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    220    228       BY SIMILARITY.
FT   DISULFID    224    236       BY SIMILARITY.
FT   DISULFID    237    245       BY SIMILARITY.
FT   DISULFID    241    253       BY SIMILARITY.
FT   DISULFID    256    265       BY SIMILARITY.
FT   DISULFID    269    296       BY SIMILARITY.
FT   DISULFID    300    311       BY SIMILARITY.
FT   DISULFID    315    330       BY SIMILARITY.
FT   DISULFID    333    337       BY SIMILARITY.
FT   DISULFID    504    513       BY SIMILARITY.
FT   DISULFID    508    521       BY SIMILARITY.
FT   DISULFID    524    533       BY SIMILARITY.
FT   DISULFID    537    553       BY SIMILARITY.
FT   DISULFID    556    569       BY SIMILARITY.
FT   DISULFID    560    577       BY SIMILARITY.
FT   DISULFID    593    615       BY SIMILARITY.
FT   DISULFID    618    626       BY SIMILARITY.
FT   DISULFID    622    634       BY SIMILARITY.
FT   CARBOHYD    114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1167 AA;  129934 MW;  4793E0749DC1D55A CRC64;

  Query Match             38.7%;  Score 2643.5;  DB 1;  Length 1167;
  Best Local Similarity   44.8%;  Pred. No. 2e-132;
  Matches  569;  Conservative  165;  Mismatches  393;  Indels  143;  Gaps   30;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              ||| :  ||| |        ||     :|| ||  :: :    : |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||    |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: | |: |
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 | ::|        ||::||| :||||| ||| :  || ||  :|| | ||  | :
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
               : ||      | |  |    |     ||        ||  |: :|| |  || || | |||:|
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
               || |||||: :|||||  ||  | |:  ||     |:  : : : ||  :|||||:||
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
               || ||: |: |:|    |      ||  |:| : |: |    | :|| |:|:  : :|:  :
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
               | :  |: | || | :    ||:|||      : || |    |::|||||| |  ||:
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
               ::  |||||||:|||||   :  :| : :|   : |||||||:|: :|  :: :   | :
Db        417 NMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRY 476

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
               :|: | :|||:  |:: : |          |   |:  |:   |||||| ||:|    |
Db        477 ANTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCVSCLHVDR 529

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFC 596
               |  ||  | :||| |||       |: || ||  |  |:||:||     |    ||::| | |
Db        530 GGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKSAHFQDGPQC 589

Qy        597 VARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAE--QRASPLTSI 654
```

```
Db        590 IPRCPHGILGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGDIVSHSSLAVGL 648

Qy        655 FNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRI 714

Db        649 VSGLLITVIVALLIVV---LLRRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQAFLRI 704

Qy        715 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 774

Db        705 LKETEFKKDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQEVLDEAYVM 764

Qy        775 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMS 834

Db        765 ASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCVQIAKGMN 824

Qy        835 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI 894

Db        825 YLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPIKWMALESI 884

Qy        895 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 954

Db        885 LQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPICTIEVYMI 944

Qy        955 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1014

Db        945 ILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSLSDRRLFSRLLSSDD- 1000

Qy       1015 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1074

Db       1001 -DVVDADEYLLPYKRI---------------------NRQGS---------------EP 1022

Qy       1075 LAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDGYVAPLTC 1133

Db       1023 CIPPTGH--------------------PVRENSITLRNISDPTQNALEKDLDGH------ 1056

Qy       1134 SPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKTLSPGKNGV 1180

Db       1057 ----EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYLNTNQNSL 1112

Qy       1181 VKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGT 1240

Db       1113 ---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGALTGNGMFL 1147

Qy       1241 PTAENPEYLG 1250

Db       1148 PAAENLEYLG 1157
```

```
RESULT 9
ERB3_HUMAN
ID   ERB3_HUMAN     STANDARD;      PRT;  1342 AA.
AC   P21860;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN   ERBB3 OR HER3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90083234; PubMed=2687875;
RA   Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT   "Isolation and characterization of ERBB3, a third member of the
RT   ERBB/epidermal growth factor receptor family: evidence for
RT   overexpression in a subset of human mammary tumors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90311312; PubMed=2164210;
RA   Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA   Todaro G.J., Shoyab M.;
RT   "Molecular cloning and expression of an additional epidermal growth
RT   factor receptor-related gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN   [3]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RC   TISSUE=Placenta;
RX   MEDLINE=93282822; PubMed=7685162;
RA   Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT   "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT   tyrosine kinase.";
RL   Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC       SECRETED (SHORT FORM).
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC       SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC   -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M29366; AAA35790.1; -.
DR   EMBL; M34309; AAA35979.1; -.
DR   EMBL; S61953; AAB26935.1; -.
DR   PIR; A36223; A36223.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3431; ERBB3.
DR   MIM; 190151; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1342       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN      20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   644    664       POTENTIAL.
FT   DOMAIN     665   1342       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     709    966       PROTEIN KINASE.
FT   NP_BIND    715    723       ATP (BY SIMILARITY).
FT   BINDING    742    742       ATP (BY SIMILARITY).
FT   ACT_SITE   834    834       BY SIMILARITY.
FT   DISULFID   186    194       BY SIMILARITY.
FT   DISULFID   190    202       BY SIMILARITY.
```

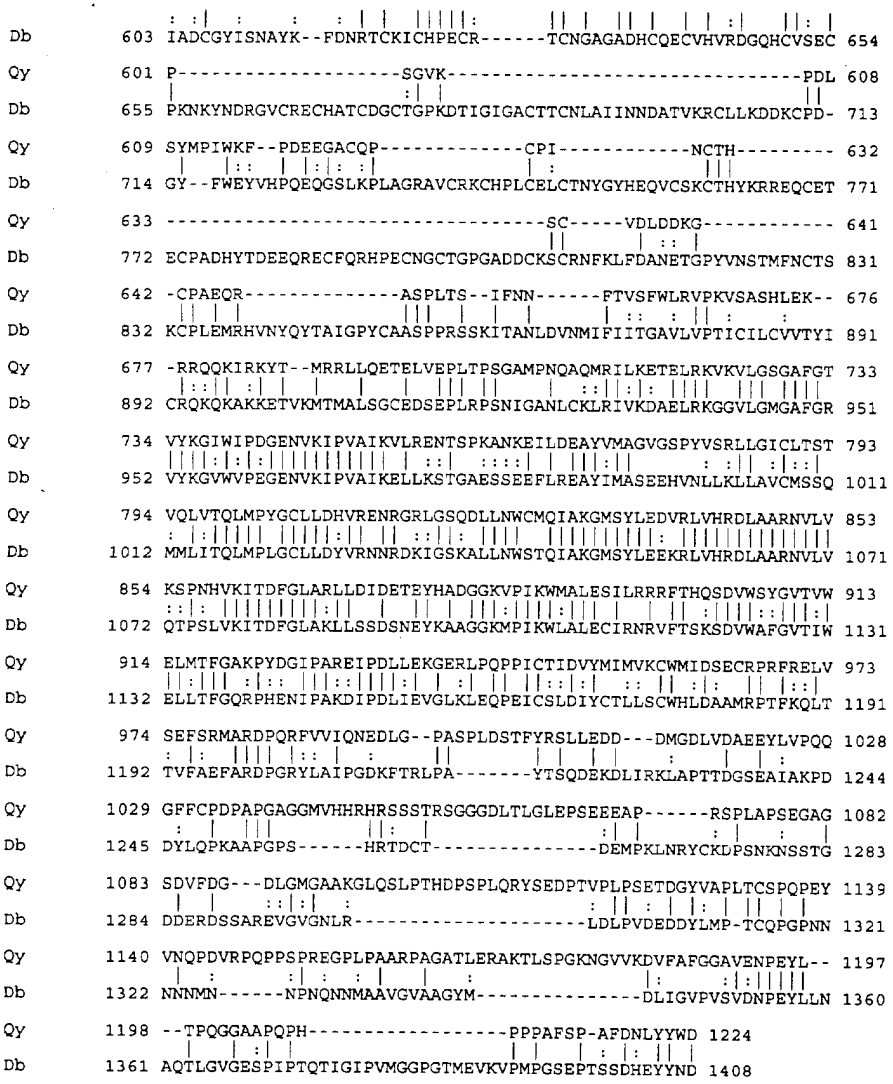
```
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    552    565       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    141    183       EILSGGVYIEKNDKLCHMDTIDWRDIVRDRDAEIVVKDNGR
FT                                SC -> GQFPMVPSGLTPQPAQDWYLLDDDPRLLTLSASSK
FT                                VPVTLAAV (IN SHORT ISOFORM).
FT   VARSPLIC    184   1342       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    560    560       E -> G (IN REF. 2).
FT   CONFLICT   1064   1064       E -> G (IN REF. 2).
SQ   SEQUENCE   1342 AA;  148097 MW;  7201E7F66CA374BD CRC64;

  Query Match           35.1%;  Score 2398.5;  DB 1;  Length 1342;
  Best Local Similarity  40.0%;  Pred. No. 2e-119;
  Matches  527;  Conservative  192;  Mismatches  458;  Indels  139;  Gaps   33;

Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
              ||| :|        |  || ||    | :     |      |  ||: |:|| ||||:
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
              || ||||| |:|| ||||:| |:    :||  ||:|||||::: :|: |:     ||  |
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
                |   ||:|:|  ||||| ||| |::| :||: ||| |:||    :     ::  :
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
              |:| ||  :||| |||| |||||:||:|:||  |   | || |  |||::|| ||:|
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
              |: :|| || ||| || |   ||  : ||  ||:  |||  :| :|  || :||:|::
Db        238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFQLEPNPHTKYQYGGVCVASCPHNFV-V 296

Qy        307 DVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF--IGITELE- 363
              |  ||   ||     ||  ::| : || |   | : | | |    : |:|  :  : ::
Db        297 DQTSCVRACPPDKMEVD-KNGLKMCEPCGGLCPKACEGTG-----SGSRFQTVDSSNIDG 350

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||    :|||       | ||:| || |: |||||| | :||  : : |
Db        351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVFRTVREITGYLNIQSWPPHMHNFS 410

Qy        424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
              || ||  | || |:|  :| | :: | :: || |||:|: :|   |  |  ||: |:: |
Db        411 VFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNW 470

Qy        483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECV 541
              ::  | | :  |    |||  :|| ||  |  ||: | |||||| ||::|  : ||  ||
Db        471 TKVLRGPTEERLDIKHNRPRRDCVAEGKVCDPLCSSGGCWGPGPGQCLSCRNYSRGGVCV 530

Qy        542 EECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCP 601
                |  | | |||: :   |  |||||||  |: || |  :| |  |||::| | ||: ||
Db        531 THCNFLNGEPREFAHEAECFSCHPECQPMEGTATCNGSGSDTCAQCAHFRDGPHCVSSCP 590

Qy        602 SGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAE-------QRASPLTSI 654
              ||    :  ||:|:|| :  |:||  |||  |   : : |  :          :  ::
Db        591 HGVLG--AKGPIYKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLIGKTHLTMALTV 648

Qy        655 FNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRI 714
                  | | :    :  : |  | ::   |  ||| |:  | :||| ||    |:  ||
Db        649 IAGLVVIFMM----LGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVLARI 703

Qy        715 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 774
              |||||||:|||||| ||||:||:|||:||::|||| |||: : :  :: : : |    :
Db        704 FKETELRKLKVLGSGVFGTVHKGVWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAI 763

Qy        775 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMS 834
              :   :: ||||:|  |::||||| :| | ||||||::|| || | |||| :||||||
Db        764 GSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVRQHRGALGPQLLLNWGVQIAKGMY 823

Qy        835 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI 894
              |||:  :|||:|||||||:|||: |:: |||:|  ||  |: :       | ||||||||||
Db        824 YLEEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKWMALESI 883

Qy        895 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 954
                ::|||||||||||||||||||:|| |:   |:||||||||| || ||||||||:
Db        884 HFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEVPDLLEKGERLAQPQICTIDVYMV 943

Qy        955 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPA---SPLDSTFYRSLLED 1011
              ||||||||   || :|| :||:|||||| |::||: |  ||     |          ||:
Db        944 MVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKLEE 1002

Qy       1012 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP-SEEEA 1070
              ::   :| :  |  ::                      : :| : |  |:| :   :
Db       1003 VELEPELDLDLDLEAEED------------------NLATTTLGSALSLPVGTLNRPRG 1043

Qy       1071 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHD-PSPLQRYSEDPTVPLP------SE 1123
              :| |:|| |    :  |:||    :   |  :   | |:  :      |:|
Db       1044 SQSLLSPSSGY-MPMNQGNLGESCQESAVSGSSERCPRPVSLH------PMPRGCLASES 1096

Qy       1124 TDGYVA----------------PLTCSPQPE----YVNQPDVRPQPPSPREGP------- 1156
                :|:|                : ||:|    | :|    | :|   |
Db       1097 SEGHVTGSEAELQEKVSMCRSRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPPGLEEEDV 1156

Qy       1157 ----LPAARPAGATLERAKTLSP-GKNGVV------KDVFAFGGAVENPEYLTPQGGAAP 1205
                :|    |   |  |||  | : |:      :|       |  ||:  :  :|
Db       1157 NGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEEEDED--------EEYEYMNRRRRHSP 1208

Qy       1206 QPHPPPAFSPAFDNLYYWD-----------QDPPERGAPPSTFKGTPTAENPEYL 1249
              ||||   |       | |            |  |    |    ||   |: ||:
Db       1209 -PHPPRPSSLEELGYEYMDVGSDLSASLGSTQSCPLHPVPIMPTAGTTPDEDYEYM 1263


RESULT 10
ERB3_RAT
ID   ERB3_RAT       STANDARD;      PRT;  1339 AA.
AC   Q62799; Q62955;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3).
GN   ERBB3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96096535; PubMed=8522190;
RA   Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT   "Cloning of the rat ErbB3 cDNA and characterization of the
RT   recombinant protein.";
RL   Gene 165:279-284(1995).
RN   [2]
RP   REVISIONS TO 85; 513 AND 565.
RA   Hellyer N.J., Koland J.G.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 922-1097 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29339; AAC28498.2; -.
DR   EMBL; U52530; AAC53050.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1339       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN      20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   644    662       POTENTIAL.
FT   DOMAIN     663   1339       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     183    259       CYS-RICH.
FT   DOMAIN     707    964       PROTEIN KINASE.
FT   NP_BIND    713    721       ATP (BY SIMILARITY).
FT   BINDING    740    740       ATP (BY SIMILARITY).
FT   ACT_SITE   832    832       BY SIMILARITY.
FT   DISULFID   186    194       BY SIMILARITY.
FT   DISULFID   190    202       BY SIMILARITY.
FT   DISULFID   210    218       BY SIMILARITY.
FT   DISULFID   214    226       BY SIMILARITY.
FT   DISULFID   227    235       BY SIMILARITY.
FT   DISULFID   231    243       BY SIMILARITY.
FT   DISULFID   246    255       BY SIMILARITY.
FT   DISULFID   259    286       BY SIMILARITY.
FT   DISULFID   290    301       BY SIMILARITY.
FT   DISULFID   305    320       BY SIMILARITY.
FT   DISULFID   323    327       BY SIMILARITY.
FT   DISULFID   500    509       BY SIMILARITY.
FT   DISULFID   504    517       BY SIMILARITY.
FT   DISULFID   520    529       BY SIMILARITY.
FT   DISULFID   533    549       BY SIMILARITY.
FT   DISULFID   556    573       BY SIMILARITY.
FT   DISULFID   576    585       BY SIMILARITY.
FT   DISULFID   589    610       BY SIMILARITY.
FT   DISULFID   613    621       BY SIMILARITY.
FT   DISULFID   617    629       BY SIMILARITY.
FT   CARBOHYD   126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT  1028   1028       L -> P (IN REF. 3).
SQ   SEQUENCE  1339 AA;  147545 MW;  0AA5F2402BBFDF1E CRC64;

  Query Match             34.0%;  Score 2325.5;  DB 1;  Length 1339;
  Best Local Similarity   40.5%;  Pred. No. 1.5e-115;
  Matches  521;  Conservative  169;  Mismatches  436;  Indels  159;  Gaps   34;

Qy          3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              |  ||      |  |  |:    |  || ||   | :    :      |  ||: |:|| ||
Db          7 LQVLC----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLYKLYEKCEVVMGN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:     || ||||| |:|| ||||:| |:   :||  ||:|||||::  :|: |:
Db         63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLPNLRVVRGTQVYDGKFAIFVM-- 120

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 ||  |     |  ||:|:  ||||| ||| |::| :||: ||| |:||
Db        121 ---LNYNT----NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRVR--- 170

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
              ::  |    | ||  :||| ||||    :||| ||:|:||  |  || || |  |||:
Db        171 GAEIVVKNNGANCPPCHEVCKG-RCWGPGPDDCQILTKTICAPQCNGRCFGPNPNQCCHD 229

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              :|| ||:||: :|| ||  || || |   ||  : ||  ||:  |||  :| :|  || :
Db        230 ECAGGCSGPQDTDCFACRRFNDSGACVPRCPEPLVYNKLTFQLEPNPHTKYQYGGVCVAS 289

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGL--GMQYIKANSKF 356
              ||:|::  |    |   ||    ||  : | : || |   | : | |   | :|   :|
Db        290 CPHNFV-VDQTFCVRACPPDKMEVD-KHGLKMCEPCGGLCPKACEGTGSGSRYQTVDSSN 347

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
              |      |  | || |:| ||    : ||       | ||:| || |: |||||  | :||
Db        348 ID----GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLNVFRTVREITGYLNIQSWP 403

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLC 475
              : : ||| ||  | || |:|  :| | :: | :: || |||:|: :|   |  |  ||
Db        404 PHMHNFSVFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLC 463

Qy        476 FVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQF 534
              : |:: | :| | | :  |     :||  ||: ||  |  ||: | |||||| ||::|  :
Db        464 YHHSLNWTRLLRGPSEERLDIKYDRPLGECLAEGKVCDPLCSSGGCWGPGPGQCLSCRNY 523

Qy        535 LRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPP 594
              |   ||  |  ||| |||:|:   |  ||||| |  |: || |  :| |  |||::| |
Db        524 SREGVCVTHCNFLQGEPREFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAHFRDGP 583

Qy        595 FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTS- 653
```

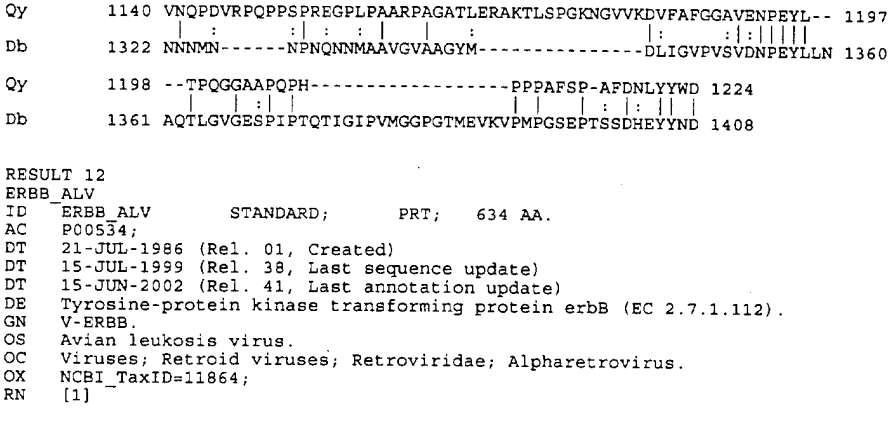
```
Db        584 HCVNSCPHGILG--AKGPIYKYPDAQNECRPCHENCTQGCNGPELQDCLGQ--AEVLMSK 639

Qy        654 --IFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQ 711

Db        640 PHLVIAVTVGLAVILMILGGSFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVL 698

Qy        712 MRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 771

Db        699 ARIFKETELRKLKVLGSGVFGTVHKGIWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHM 758

Qy        772 YVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAK 831

Db        759 LAVGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVKQHRETLGPQLLLNWGVQIAK 818

Qy        832 GMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMAL 891

Db        819 GMYYLEEHSMVHRDLALRNVMLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKWMAL 878

Qy        892 ESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDV 951

Db        879 ESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPQICTIDV 938

Qy        952 YMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLED 1011

Db        939 YMVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAEPSVLTT 995

Qy       1012 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE--- 1068

Db        996 KEL-----QEAELEPEL----------------------------DLDLDLEAEEEGLA 1021

Qy       1069 -----------------EAPRSPLAPSEG--------AGSDVFDGDLGMGAAKGLQSLPT 1103

Db       1022 TSLGSALSLPTGTLTRPRGSQSLLSPSSGYMPMNQSSLGEACLDSAVLGGREQFSRPISL 1081

Qy       1104 HDPSPLQRYSEDPTVPLPSETDGYV----APL-----TC-------SPQPE----YVNQP 1143

Db       1082 H-PIPRGR-------PASESSEGHVTGSEAELQEKVSVCRSRSRSRSPRPRGDSAYHSQR 1133

Qy       1144 DVRPQPPSPREGP-----------LPAARPAGATLERAKTLSP-GKNGVV------KDVF 1185

Db       1134 HSLLTPVTPLSPPGLEEEDGNGYVMPDTHLRGASSSREGTLSSVGLSSVLGTEEEDED-- 1191

Qy       1186 AFGGAVENPEYLTPQGGAAPQPHPP 1210

Db       1192 ------EEYEYMNRKRRGSP-PRPP 1209
```

RESULT 11
EGFR_DROME
ID   EGFR_DROME   STANDARD;   PRT; 1426 AA.
AC   P04412; O61601; Q9W2G0; P81868;
DT   13-AUG-1987 (Rel. 05, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE   (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN   EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX   MEDLINE=94350209; PubMed=8070664;
RA   Clifford R., Schupbach T.;
RT   "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT   that several genetically defined classes of alleles cluster in
RT   subdomains of the receptor protein.";
RL   Genetics 137:531-550(1994).
RN   [2]
RP   REVISIONS.
RA   Clifford R., Schupbach T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=85124611; PubMed=2982499;
RA   Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT   "The Drosophila EGF receptor gene homolog: conservation of both
RT   hormone binding and kinase domains.";
RL   Cell 40:599-607(1985).
RN   [4]
RP   SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC   STRAIN=Oregon-R; TISSUE=Embryo;
RX   MEDLINE=87002474; PubMed=3093080;
RA   Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT   "Alternative 5' exons and tissue-specific expression of the
RT   Drosophila EGF receptor homolog transcripts.";
RL   Cell 46:1091-1101(1986).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=99102120; PubMed=9882502;
RA   Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT   "Several levels of EGF receptor signaling during photoreceptor
RT   specification in wild-type, Ellipse, and null mutant Drosophila.";
RL   Dev. Biol. 205:129-144(1999).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [7]
RP   SEQUENCE OF 959-1078 FROM N.A.
RC   STRAIN=Daekwanryeong;
RX   MEDLINE=85137938; PubMed=2983232;
RA   Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT   "A Drosophila genomic sequence with homology to human epidermal

```
RT   growth factor receptor.";
RL   Nature 314:178-180(1985).
RN   [8]
RP   SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=92038942; PubMed=1936959;
RA   Raz E., Schejter E.D., Shilo B.Z.;
RT   "Interallelic complementation among DER/flb alleles: implications for
RT   the mechanism of signal transduction by receptor-tyrosine kinases.";
RL   Genetics 129:191-201(1991).
RN   [9]
RP   REVIEW.
RX   MEDLINE=97248481; PubMed=9094709;
RA   Perrimon N., Perkins L.A.;
RT   "There must be 50 ways to rule the signal: the case of the Drosophila
RT   EGF receptor.";
RL   Cell 89:13-16(1997).
CC   -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC       WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC       MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC       CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC       DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC       POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC       ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC       AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC       SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC       CUTICLE.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC       MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC       PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC       TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC       UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC       TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC       EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC       IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC       POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC       AND THORACIC AND ABDOMINAL GANGLIA.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF052754; AAC08536.1; -.
DR   EMBL; AF052753; AAC08536.1; JOINED.
DR   EMBL; AF052754; AAC08535.1; -.
DR   EMBL; AF052752; AAC08535.1; JOINED.
DR   EMBL; K03054; AAA51462.1; -.
DR   EMBL; K03417; AAA51460.1; -.
DR   EMBL; K03416; AAA50965.1; -.
DR   EMBL; K03418; AAA51461.1; -.
DR   EMBL; AF109077; AAD26134.1; -.
DR   EMBL; AF109078; AAD26132.1; -.
DR   EMBL; AF109082; AAD26132.1; JOINED.
DR   EMBL; AF109078; AAD26133.1; -.
DR   EMBL; AF109084; AAD26133.1; JOINED.
DR   EMBL; AF109079; AAD26130.1; -.
DR   EMBL; AF109081; AAD26130.1; JOINED.
DR   EMBL; AF109079; AAD26131.1; -.
DR   EMBL; AF109083; AAD26131.1; JOINED.
DR   EMBL; AF109080; AAD26135.1; -.
DR   EMBL; AE003454; AAF46732.1; -.
DR   EMBL; X02293; CAA26157.1; -.
DR   EMBL; X78920; CAA55523.1; -.
DR   EMBL; X78918; CAA55521.1; -.
DR   EMBL; X78919; CAA55522.1; -.
DR   PIR; A00640; GQFFE.
DR   HSSP; P11362; 1FGK.
DR   FlyBase; FBgn0003731; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 7.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW   Developmental protein.
FT   SIGNAL       1     30       POTENTIAL.
FT   CHAIN       31   1426       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN      31    868       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   869    889       POTENTIAL.
FT   DOMAIN     890   1426       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     938   1198       PROTEIN KINASE.
FT   NP_BIND    944    952       ATP (BY SIMILARITY).
FT   BINDING    971    971       ATP (BY SIMILARITY).
FT   ACT_SITE  1063   1063       BY SIMILARITY.
FT   MOD_RES    902    902       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

  Query Match           28.3%;  Score 1936;  DB 1;  Length 1426;
  Best Local Similarity  32.7%;  Pred. No. 5.8e-95;
  Matches  467;  Conservative  183;  Mismatches  432;  Indels  346;  Gaps   43;

Qy         24 QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT-NASLSFLQDIQEVQG 82
              ::| ||  :| :|:: | |   ||  |  |  | |||:||:||  |  |||| :|:|| |
Db        100 KICIGTKSRLSVPSNKEHHYRNLRDRYTNCTYVDGNLKLTWLPNENLDLSFLDNIREVTG 159

Qy         83 YVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGL 137
              |:||:|  |::|   :|:|:||  ||     |: ||| |                :   :
Db        160 YILISHVDVKKVVFPKLQIIRGRTLFSLSVEEEKYALFV---------------TYSKM 203

Qy        138 RELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSP 197
              |::  | ::| | |   |  ||:  || | :|              |  | | |
Db        204 YTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPKCHE 263

Qy        198 MCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDCLAC 255
              |    |||| ::||  ::  |:  ||  || || | :|||  || |||||   ||:||
Db        264 SCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDCIAC 322

Qy        256 LHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVC 315
              :|    : :  || :  ||  |:    ||||:|:||:||  ||  :|  | |:|   |
Db        323 KNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACVRSC 381

Qy        316 PLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAGCKK 369
              |  :  |   |  |:|| : |                |:| |       | |
Db        382 PQDKMDKGGE-----CVPCNGPCPKTC-------------PGVTVLHAGNIDSFRNCTV 422

Qy        370 IFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDSLPD 421
              | |::  | ::| |  |  :|          || ||: :|| |::|||||| |      :
Db        423 IDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQFRN 482

Qy        422 LSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              || |:||: | || |     : :| :     :  | :|:|::: ||  :| ||  ||:|  :
Db        483 LSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYVSNI 542

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |  : : | | :    |  | |  |  |    |||  |  ||: |  |      |
Db        543 RWPAIQKEPEQKVWVNENLRADLCEKNGTICSDQCNEDGCWGAGTDQCLTCKNFNFNGTC 602

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
```



```
RESULT 12
ERBB_ALV
ID   ERBB_ALV       STANDARD;      PRT;   634 AA.
AC   P00534;
DT   21-JUL-1986 (Rel. 01, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian leukosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11864;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85228222; PubMed=2988784;
RA   Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA   Crittenden L.B., Raines M.A., Kung H.-J.;
RT   "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT   processing and promoter insertion result in expression of an
RT   amino-truncated EGF receptor.";
RL   Cell 41:719-726(1985).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC       IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC       AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC       THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC       C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC       PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10066; AAA48763.1; ALT_INIT.
DR   PIR; A00643; TVCHLV.
DR   PIR; B00643; TVFVLV.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
SQ   SEQUENCE   634 AA;  70891 MW;  E705E33A0BE01FCC CRC64;

  Query Match             24.9%;  Score 1704.5;  DB 1;  Length 634;
  Best Local Similarity   51.6%;  Pred. No. 4e-83;
  Matches  364;  Conservative   72;  Mismatches  149;  Indels  121;  Gaps   15;
```



```
                 ||||||||| | :|||||||||||||||||||||:|||||||| ||  :||||||||||||
Db        299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy        947 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1005
                 |||||||||||||||:: ||:||||::|||:||||| |::||| :| :   || || ||
Db        359 CTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy       1006 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1065
                |:|:|::|| |:|||:||||| ||||                  | ||
Db        419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------ 449

Qy       1066 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1120
                    |:||     | |  : ::  |        :  |  |  : | :|||| |||
Db        450 -----SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNF 496

Qy       1121 PSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKN 1178
                |:  ||::      | ||||||  : |: ||                  :  :|
Db        497 LEESIDDGFL------PAPEYVNQ--LMPKKPS--------------------TAMVQN 527

Qy       1179 GVVKDVF---------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYW 1223
                 :  ::                :  ||:|||||          :  |     |::  ||
Db        528 QIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFESSPYW 579

Qy       1224 DQ--------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1254
                |        | |:           |:     | |||||| :  |
Db        580 IQSGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 625
```

```
RESULT 13
ERBB_AVIER
ID   ERBB_AVIER     STANDARD;      PRT;   604 AA.
AC   P00535;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ES4).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=79685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H;
RX   MEDLINE=84026539; PubMed=6313229;
RA   Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT   "The erbB gene of avian erythroblastosis virus is a member of the src
RT   gene family.";
RL   Cell 35:71-78(1983).
RN   [2]
RP   SEQUENCE OF 1-152 FROM N.A.
RX   MEDLINE=84223957; PubMed=6328658;
RA   Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA   Saule S., Martin P., Stehelin D.;
RT   "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT   new type of oncogene.";
RL   Science 224:1456-1459(1984).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC       ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC       IN CHICKENS.
CC   -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC       RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; K02006; AAA42394.1; ALT_INIT.
DR   EMBL; K01216; AAA42400.1; -.
DR   PIR; A00644; TVYUH.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   CONFLICT     29     29       R -> W (IN REF. 2).
FT   CONFLICT    140    140       S -> F (IN REF. 2).
FT   CONFLICT    146    146       I -> V (IN REF. 2).
SQ   SEQUENCE   604 AA;  67633 MW;  76EBCDD06745D609 CRC64;
```

```
  Query Match             24.3%;  Score 1658;  DB 1;  Length 604;
  Best Local Similarity   51.4%;  Pred. No. 1.1e-80;
  Matches  354;  Conservative   69;  Mismatches  142;  Indels  124;  Gaps   14;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
                |||: | | ||  ||:||  :   : : |: |    || |  |||  |     :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAM 706
                  |  ||           |      |  ||:  :|| |:|||||| ||||||||||
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEA 118

Qy        707 PNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 766
                |||| :|||||||| :|||||||||||||:|||:|||:|| |||||||| ||| ||||||||
Db        119 PNQAHLRILKETEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKELREATSPKANKE 178

Qy        767 ILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWC 826
                |||||||| | :|:| |||||||||||||:|||||||||||::||::  :||| |||||
Db        179 ILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWC 238

Qy        827 MQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPI 886
                :||||||:|||: |||||||||||||||:| ||||||||||:||  || ||||:||||||
Db        239 VQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPI 298

Qy        887 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 946
                |||||||| | :|||||||||||||||||||:|||||||| ||  :|||||||||||
Db        299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy        947 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1005
                |||||||||||||||:: ||:||||::|||:||||| |::||| :| :   || || ||
Db        359 CTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy       1006 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1065
                |:|:|::|| |:|||:||||| ||||                  | ||
Db        419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------ 449

Qy       1066 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1120
                    |:||     | |  : ::  |        :  |  |  : | :|||| |||
Db        450 -----SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNF 496

Qy       1121 PSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKN 1178
                |:  ||::      | ||||||  : |: ||
Db        497 LEESIDDGFL------PAPEYVNQ--LMPKKPSTAM----------------------- 524

Qy       1179 GVVKDVFAF-------------------GGAVENPEYLTPQGGAAPQPHPPPAFSPAFDN 1219
                |   :: |                   ||:|||||          :  |     |::
Db        525 -VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFES 575

Qy       1220 LYYWDQDPPERGAPPSTFKGTPTAENPEY 1248
                || |    :             :||:|
```

```
Db        576 SPYWIQSGNHQ----------INLDNPDY 594


RESULT 14
ERBB_AVIEU
ID   ERBB_AVIEU     STANDARD;      PRT;   540 AA.
AC   P11273;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ts167).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=103898;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87064458; PubMed=2878364;
RA   Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT   "A single amino acid substitution in v-erbB confers a thermolabile
RT   phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT   cells.";
RL   Mol. Cell. Biol. 6:1751-1759(1986).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13179; AAA42401.1; -.
DR   PIR; A25231; TVFVEB.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   VARIANT     270    270       H -> D (IN THERMOLABILE V-ERBB).
SQ   SEQUENCE   540 AA;  60412 MW;  5B53297AA068B65D CRC64;

  Query Match             23.2%;  Score 1585;  DB 1;  Length 540;
  Best Local Similarity   54.2%;  Pred. No. 6.8e-77;
  Matches  335;  Conservative   62;  Mismatches  133;  Indels   88;  Gaps   12;

Qy        587 CAHYKDPPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
              |||: | | ||  ||:|| :   : :||: |    || |  |||  |     :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAM 706
                |   ||           |       |  ||:  :|| |:|||||| ||||||||||
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEA 118

Qy        707 PNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 766
              |||| :|||||| :|||||| |||||||||:|||:|| | |||||| ||| ||||||||
Db        119 PNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELREATSPKANKE 178

Qy        767 ILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWC 826
              |||||||||| | :|:| |||||||||||||:||||||||||::||::  :||| |||||

Db        179 ILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWC 238

Qy        827 MQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPI 886
              :||||||:|||:  :|||||||||||||:| ||||||||||: |  || ||||:||||||
Db        239 VQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAKQLGADEKEYHAEGGKVPI 298

Qy        887 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 946
              |||||||| | :|||||||||||||||||||:|||||||| ||  :||||||||||||
Db        299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy        947 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1005
              |||||||||||||| |:: ||:||||::|||:||||| |::||| :| :   || || ||
Db        359 CTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy       1006 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1065
              |:|:|::|| |:|||:||||| |||| | ||
Db        419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------- 449

Qy       1066 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1120
                    |:||     | |  : ::  | : |          |
Db        450 -----SRTPLLSSLSATSNNSATNCIDRNGG----------H---------------- 476

Qy       1121 PSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERAKTLSPGKNG 1179
              |   ||::      | |||||  : |: ||          | : :   |  :|
Db        477 PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAISKLPIDSRYQN- 527

Qy       1180 VVKDVFAFGGAVENPEYL 1197
                    :  ||:|||||
Db        528 ------SHSTAVDNPEYL 539


RESULT 15
EGFR_CHICK
ID   EGFR_CHICK     STANDARD;      PRT;   703 AA.
AC   P13387;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE   (Fragment).
GN   EGFR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88261272; PubMed=3260329;
RA   Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA   Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT   "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT   expression in mouse cells, and differential binding of EGF and
RT   transforming growth factor alpha.";
RL   Mol. Cell. Biol. 8:1970-1978(1988).
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20386; AAA48760.1; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 4.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     30
FT   CHAIN        31   >703       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    667       POTENTIAL.
FT   DOMAIN      668   >703       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    197    206       BY SIMILARITY.
FT   DISULFID    201    214       BY SIMILARITY.
FT   DISULFID    222    230       BY SIMILARITY.
FT   DISULFID    226    238       BY SIMILARITY.
FT   DISULFID    239    247       BY SIMILARITY.
FT   DISULFID    243    255       BY SIMILARITY.
FT   DISULFID    258    267       BY SIMILARITY.
FT   DISULFID    271    298       BY SIMILARITY.
FT   DISULFID    302    314       BY SIMILARITY.
FT   DISULFID    318    333       BY SIMILARITY.
FT   DISULFID    336    340       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    581       BY SIMILARITY.
FT   DISULFID    569    589       BY SIMILARITY.
FT   DISULFID    592    601       BY SIMILARITY.
FT   DISULFID    605    627       BY SIMILARITY.
FT   DISULFID    630    638       BY SIMILARITY.
FT   DISULFID    634    646       BY SIMILARITY.
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    200    200       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    359    359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    420    420       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    633    633       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    648    648       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     703    703
SQ   SEQUENCE   703 AA;  77427 MW;  AFF2DE11B735A690 CRC64;

  Query Match             22.7%;  Score 1550;  DB 1;  Length 703;
  Best Local Similarity   43.8%;  Pred. No. 6.4e-75;
  Matches  310;  Conservative  104;  Mismatches  265;  Indels   28;  Gaps   12;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |   :| ||  | |        :|| ||: ||      | |   |: :|  |:||  |||
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  |:||: |||| |||||| | |  :||: |:|:||  |:::::|||| |
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
              :| |          |||| :: |:||| ||| |  ||:||  ||:|| ||   : :  |
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              .|::|  :   :|  | | |    |||   ::||:||: :||  |: ||:| :|:|||| |

Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: ||||||  |      |:  || || ||  |::   ||||:|:|||:||  |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||    |      ||  |:| ::|:||   |::|| |:|:  :|     |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  | || | :: || :| ||  : | ||  |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
                ||  |:||::|||   :| ||| :  | |  ||||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: |  ||     |      || :::|  :   |  ||:  ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
              |||::| :||| |||:      ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIF 655
              ||  ||:||  :   : :||: |    || |  |||  |      :|||     |   ||
Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 NNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTP 702
                          |     | ||:  :|| |:||||| |||||||
Db        657 AGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTP 703


Search completed: July 22, 2003, 09:18:05
Job time : 21.642 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.8246 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-632-652-14
Perfect score: 6834
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

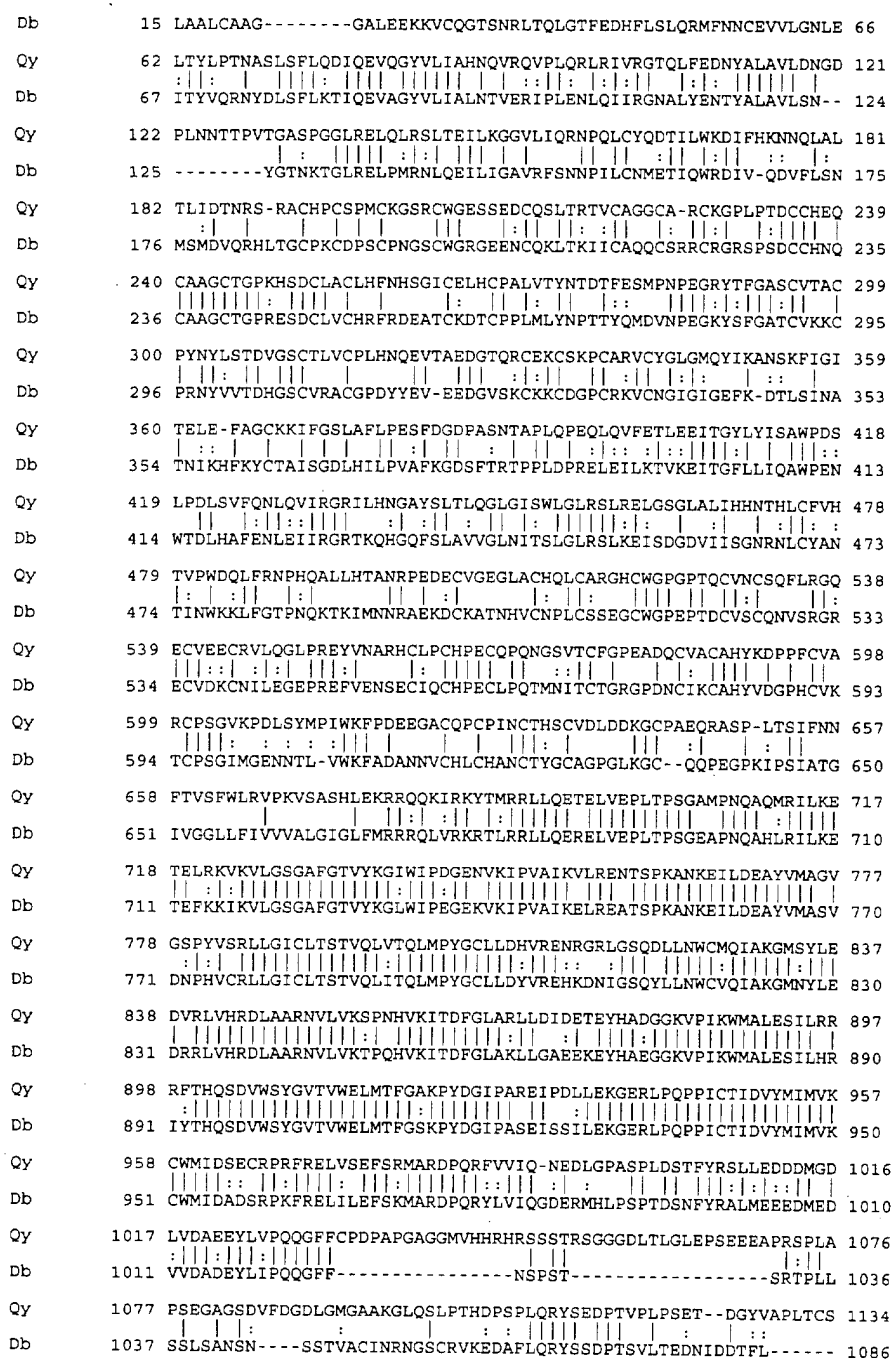
SUMMARIES

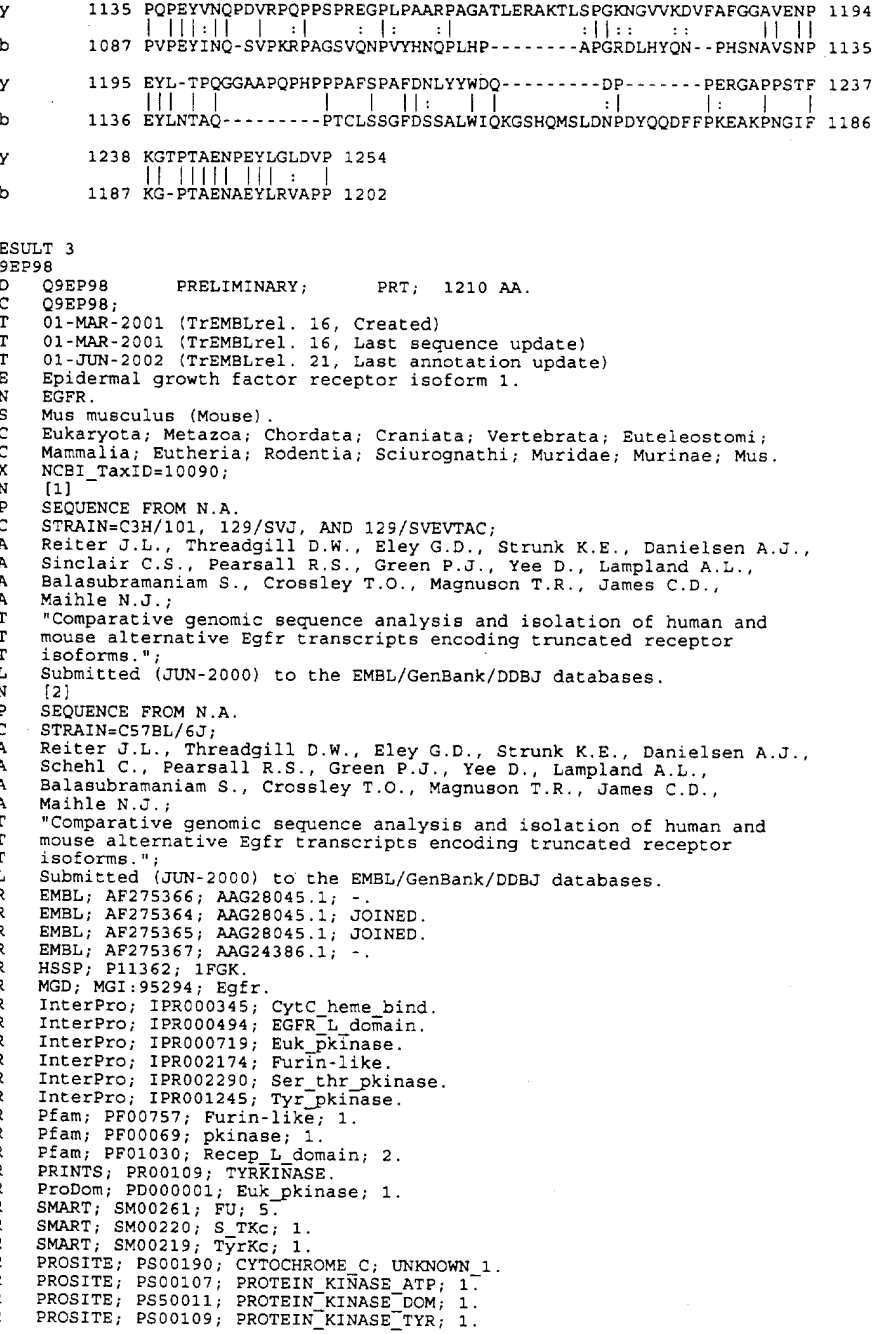
Result No.	Score	% Query Match	Length	DB	ID	Description
1	6125	89.6	1259	6	O18735	O18735 canis famil
2	3107	45.5	1209	11	Q9QX70	Q9qx70 rattus norv
3	3078	45.0	1210	11	Q9EP98	Q9ep98 mus musculu
4	2685	39.3	1165	13	Q9YH40	Q9yh40 xiphophorus
5	2646.5	38.7	1137	13	Q9W6F6	Q9w6f6 gallus gall
6	2260	33.1	1328	13	P79754	P79754 fugu rubrip
7	1988.5	29.1	1433	5	Q9BIH9	Q9bih9 anopheles g
8	1871	27.4	419	4	Q9UK79	Q9uk79 homo sapien
9	1739	25.4	367	11	Q8R2X1	Q8r2x1 mus musculu
10	1697.5	24.8	412	4	Q8WYV0	Q8wyv0 homo sapien
11	1675	24.5	729	15	Q86712	Q86712 avian rous-
12	1673	24.5	567	15	Q86714	Q86714 avian rous-
13	1608.5	23.5	962	15	Q64895	Q64895 avian eryth
14	1600	23.4	545	15	Q85468	Q85468 avian eryth
15	1506.5	22.0	655	11	Q9WVF5	Q9wvf5 mus musculu
16	1490.5	21.8	643	11	Q9ERV6	Q9erv6 mus musculu
17	1247	18.2	1193	5	Q9Y1X8	Q9y1x8 ephydatia f
18	1181.5	17.3	1368	5	Q23821	Q23821 caenorhabdi
19	1150	16.8	1717	5	Q26566	Q26566 schistosoma
20	1126	16.5	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11	Q9ESE0	Q9ese0 rattus norv
22	942.5	13.8	599	13	Q9PSH2	Q9psh2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	887	13.0	176	11	Q923V5	Q923v5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8szw1 drosophila
27	754.5	11.0	311	13	Q99162	Q99162 xiphophorus
28	739.5	10.8	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
29	734	10.7	331	4	Q9BUD7	Q9bud7 homo sapien
30	723	10.6	149	6	Q9BG66	Q9bg66 oryctolagus
31	715	10.5	1671	5	Q9NJV5	Q9njv5 biomphalari
32	688	10.1	1418	13	O93457	O93457 scophthalmu
33	675.5	9.9	1368	13	Q8UW85	Q8uw85 paralichthy
34	673.5	9.9	1369	13	Q8UW86	Q8uw86 paralichthy
35	657.5	9.6	1472	5	Q9U5A8	Q9u5a8 bombyx mori
36	657	9.6	1358	13	O73798	O73798 xenopus lae
37	652	9.5	1412	13	Q8UW84	Q8uw84 paralichthy
38	635.5	9.3	1418	13	Q8UW83	Q8uw83 paralichthy
39	631	9.2	1245	13	Q9YGH8	Q9ygh8 scophthalmu
40	629	9.2	1371	11	Q9QVW4	Q9qvw4 rattus sp.
41	621.5	9.1	2144	5	Q9VD94	Q9vd94 drosophila
42	589.5	8.6	1091	4	Q9UMQ4	Q9umq4 homo sapien
43	587.5	8.6	1036	4	Q07912	Q07912 homo sapien
44	586.5	8.6	1055	11	O54967	O54967 mus musculu
45	580.5	8.5	1072	4	Q9BTB0	Q9btb0 homo sapien

ALIGNMENTS

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErbB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB008451; BAA23127.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

```
  Query Match           89.6%;  Score 6125;  DB 6;  Length 1259;
  Best Local Similarity  89.8%;  Pred. No. 0;
  Matches 1132;  Conservative   40;  Mismatches   82;  Indels    6;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||||||||||| ||| |||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGLLLALLPSGAAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| |||||||||||||||||||:||||:|||||||||||||||||||||||||||
Db         61 ELTYLPANASLSFLQDIQEVQGYVLIAHSQVRQIPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||     |  ||: |||||||||||||||||||||||:||||:|||||||:||||||||
Db        121 DPLEGGIPAPGAAQGGLRELQLRSLTEILKGGVLIQRSPQLCHQDTILWKDVFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||  || |||| || : ||| || ||||||||||||||||||| |||||||||
Db        181 LTLIDTNRFSACPPCSPACKDAHCWGASSGDCQSLTRTVCAGGCARCKGPQPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||:|||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                |||||||||||||||||||||||||||||||:||| ||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLRVFEALEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              :|||||:|||||:||:|||||||||||||||||||||||||||||||| |  |||||||
Db        421 NLSVFQNLRVIRGRVLHDGAYSLTLQGLGISWLGLRSLRELGSGLALIHRNARLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||:|||||:|||||||||:  || |||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHSANRPEEECVGEGLACYP-CAHGHCWGPGPTQCVNCSQFLRGQEC 539

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |||||||||||||  |:||||| |||||||||||| ||||||||||||||||||||||
Db        540 VEECRVLQGLPREYVKDRYCLPCHSECQPQNGSVTCFGSEADQCVACAHYKDPPFCVARC 599

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||:|||||| |||| |||||||||||| |||:||||||||||:|||
Db        600 PSGVKPDLSFMPIWKFADEEGTCQPCPINCTHSCADLDEKGCPAEQRASPVTSIIAAVVG 659

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                  |  :    | |||:|||||||||||||||||||||||||||||||||||||||
Db        660 ILLAVVVGLVLGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 719

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 779

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              |||||||||||||||||||||||||||||:|||||||||||||:|||||||||||||
Db        780 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYLEDVR 839

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db        840 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIPPRRFT 899

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 959

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db        960 IDSECRPRFRELVAEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1019

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              |||||||||||||:| |||||  | ||||||||:|||:|||||||||||| |:||||||||
Db       1020 EEYLVPQQGFFCPEPTPGAGGTAHRRHRSSSTRNGGGELTLGLEPSEEEPPKSPLAPSEG 1079

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              |||||||||||||||||||||||: |||||||||||||||| |||| ||||||||||||
Db       1080 AGSDVFDGDLGMGAAKGLQSLPSQDPSPLQRYSEDPTVPLPPETDGKVAPLTCSPQPEYV 1139

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLER-----AKTLSPGKNGVVKDVFAFGGAVENPE 1195
              |||:| ||||   ||||| :|||||||||      ||||||||||||||||| ||||||
Db       1140 NQPEVWPQPPLALEGPLPPSRPAGATLERPKTLSPKTLSPGKNGVVKDVFAFGSAVENPE 1199

Qy       1196 YLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              || |:| |||||||||||||||||||||||| |||:|||||:||||||||||||||||
Db       1200 YLAPRGRAAPQPHPPPAFSPAFDNLYYWDQDPSERGSPPSTFEGTPTAENPEYLGLDVPV 1259


RESULT 2
Q9QX70
ID   Q9QX70       PRELIMINARY;      PRT;  1209 AA.
AC   Q9QX70;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor.
GN   EGFR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RX   MEDLINE=90258888; PubMed=2342466;
RA   Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA   Earp H.S.;
RT   "A truncated, secreted form of the epidermal growth factor receptor is
RT   encoded by an alternatively spliced transcript in normal rat tissue.";
RL   Mol. Cell. Biol. 10:2973-2982(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Petch L.A.;
RL   Submitted (NOV-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Guttridge K., Dawson T.L., Earp H.S.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M37394; AAF14008.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1209 AA;  134891 MW;  96FEE7F6CC1B7773 CRC64;

  Query Match           45.5%;  Score 3107;  DB 11;  Length 1209;
  Best Local Similarity  49.8%;  Pred. No. 1.8e-228;
  Matches  636;  Conservative  161;  Mismatches  366;  Indels  114;  Gaps   26;

Qy          3 LAALCRWGLLLALLPPGA-ASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |||||  |          ||    :|| ||  :|    : | |   |: ::  |:|| ||||
```



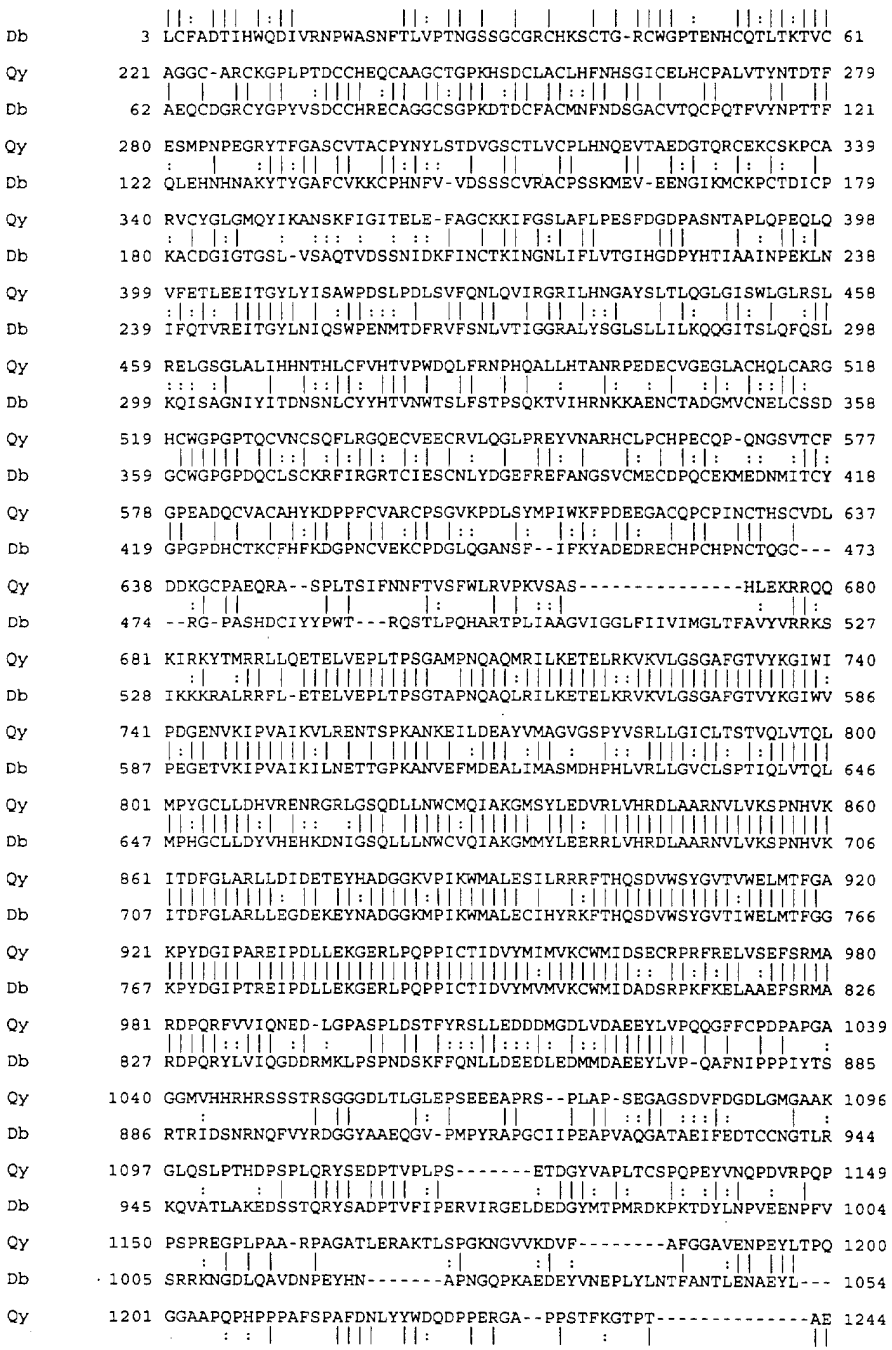


RESULT 3
Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maihle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maihle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```
KW   ATP-binding; Receptor; Transferase.
SQ   SEQUENCE   1210 AA;  134840 MW;  62CD021C9DE32E18 CRC64;

  Query Match             45.0%;  Score 3078;  DB 11;  Length 1210;
  Best Local Similarity   49.3%;  Pred. No. 3e-226;
  Matches  627;  Conservative  163;  Mismatches  372;  Indels  110;  Gaps   23;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSF 662
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 WLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRK 722
Db        657 LFIVVVALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFKK 716

Qy        723 VKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYV 782
Db        717 IKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHV 776

Qy        783 SRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLV 842
Db        777 CRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLV 836

Qy        843 HRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQ 902
Db        837 HRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQ 896

Qy        903 SDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMID 962
Db        897 SDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVKCWMID 956

Qy        963 SECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAE 1021
Db        957 ADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVVDAD 1016

Qy       1022 EYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGA 1081
Db       1017 EYLTPQQGFF----------------NSPST-----------------SRTPLLSSLSA 1042

Qy       1082 GSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEY 1139
Db       1043 TSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------PVPEY 1092

Qy       1140 VNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-T 1198
Db       1093 VNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNPEYLNT 1141

Qy       1199 PQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTPT 1242
Db       1142 AQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIFKG-PT 1191

Qy       1243 AENPEYLGLDVP 1254
Db       1192 AENAEYLRVAPP 1203
```

```
RESULT 4
Q9YH40
ID   Q9YH40      PRELIMINARY;      PRT;  1165 AA.
AC   Q9YH40;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase proto-oncogene.
GN   XMRK.
OS   Xiphophorus xiphidium.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8086;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RX   MEDLINE=98241172; PubMed=9582016;
RA   Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA   Altschmied J., Schartl M.;
RT   "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT   overexpression and mutational alterations.";
RL   Oncogene 16:1681-1690(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53471; AAD10500.2; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
```

```
SQ   SEQUENCE   1165 AA;  129614 MW;  7F7EE38D8771A74E CRC64;

  Query Match             39.3%;  Score 2685;  DB 13;  Length 1165;
  Best Local Similarity   45.3%;  Pred. No. 3.2e-196;
  Matches  579;  Conservative  161;  Mismatches  386;  Indels  152;  Gaps   32;

Qy          1 MELAALCRWGLLLALLPPG-AAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQV 55
              :||  |    ||| ||  |   ||    :|| ||  :: :    : |   :: :| || |
Db          4 LELLEL----LLLLLLSIGRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNV 56

Qy         56 VQGNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALA 115
              |  |||:||   |  |||| |||| |||||  |:|  :|| ||::||  |:| |: |
Db         57 VLENLEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLL 116

Qy        116 VLDNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHK 175
              |: |   | ::|       ||::||| :|||| ||| :  || ||  :|| | ||  |
Db        117 VMSNYQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDK 173

Qy        176 NNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTD 234
              :  : ||      | | | | |    ||       ||  |: :||  |  ||:|| | |
Db        174 TSNPTMNLIPHAFERQCQKCDPGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPID 233

Qy        235 CCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
              ||:| || ||||||: :|||||  ||  | |:  ||     |:  : : : ||  :|||||:
Db        234 CCNEHCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAA 293

Qy        295 CVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGM----QYI 350
              ||  || ||: |: |:|   |    ||  |:| : |: |   | :|| |:|:      |
Db        294 CVKECPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSLSNTI 351

Qy        351 KANSKFIGITELEPAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
                ||  ||     |: | || | :    ||:|||     |: || |    |::||||||
Db        352 AVNSTNIG----SFSNCTKINGDIILNRNSFEGDPHYKIGPMDPEHLWNLTTVKEITGYL 407

Qy        411 YISAWPDSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIH 469
              |  ||:::  |||||||::||||   :  :| : :|   : |||| ||:|: :|  ::
Db        408 VIMWWPENMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVSHLQWLGLSSLKEVSAGNVILK 467

Qy        470 HNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCV 529
              :   | :  |: | :|||:  |:: : |         |    |:  |:  |||||| ||
Db        468 NTPQLRYASTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCV 520

Qy        530 NCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAH 589
              :|    ||  ||  | :||| |||       |: || ||  |  |:||:||    |  |||
Db        521 SCLHVDRGGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKCAH 580

Qy        590 YKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAE--QR 647
              ::| | |: ||| |:  |   : |||: |: | ||||  |||  |      ||  :
Db        581 FQDGPQCIPRCPHGMLGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGDIVSH 639

Qy        648 ASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMP 707
              :|    :  :   ::  : :  |    | :||: | || |:|||||| ||||||||||  |
Db        640 SSLAVGLVSGLLITVIVALLIVV---LLRRRRIK-RKRTIRRLLQEKELVEPLTPSGQAP 695

Qy        708 NQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEI 767
              ||| :|||||| :| :|||||||||||||:| ||||::||||||||| |||| |:|:
Db        696 NQAFLRILKETEFKKDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQEV 755

Qy        768 LDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCM 827
              |||||||  |  |:| |||||||| |||||||||||||:||::: |:  | |||||:
Db        756 LDEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCV 815

Qy        828 QIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIK 887
              ||||||:|||:  |||||||||||:|:||||||||||::||  || || | |||||||
Db        816 QIAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQAHGGKVPIK 875

Qy        888 WMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPIC 947
              ||||||||:  :|||||||||||||||||||:|||||||:||  :|| ||||||||||
Db        876 WMALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPIC 935

Qy        948 TIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRS 1007
              ||:||||::|||||   |||||||| |||:||||| |::|||       || |   :
Db        936 TIEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSPSDRRLFSR 992

Qy       1008 LLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSE 1067
              ||  ||  |:|||:|||:                      |::  : | |
Db        993 LLSSDD--DVVDADEYLL--------------------RYKRIN-RQGS---------- 1018

Qy       1068 EEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDG 1126
                    | | |                                | : |   ||  |||   |  : ||
Db       1019 -----EPCIPPNGH--------------------PVRENSIALRYISDPTQNALEKDLDG 1053

Qy       1127 YVAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKTL 1173
              :      ||||||              |:  :    ||:  ::: |     | : |
Db       1054 H----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYL 1103

Qy       1174 SPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAP 1233
              :  :| :        |::::|:|   | |          :  ||          |: ||
Db       1104 NTNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGAL 1138

Qy       1234 PSTFKGTPTAENPEYLGL 1251
                    | ||| |||||
Db       1139 TGNGMFLPAAENLEYLGL 1156


RESULT 5
Q9W6F6
ID   Q9W6F6      PRELIMINARY;      PRT;  1137 AA.
AC   Q9W6F6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase (Fragment).
GN   ERBB4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HINDBRAIN;
RX   MEDLINE=99263203; PubMed=10328884;
RA   Dixon M., Lumsden A.;
RT   "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
RT   embryonic chick hindbrain.";
RL   Mol. Cell. Neurosci. 13:237-258(1999).
DR   EMBL; AF121963; AAD31764.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001368; TNFR_c6.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 1.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW   Kinase; Tyrosine-protein kinase.
FT   NON_TER       1      1
SQ   SEQUENCE   1137 AA;  127927 MW;  4D616436F87DC84F CRC64;

  Query Match             38.7%;  Score 2646.5;  DB 13;  Length 1137;
  Best Local Similarity   46.5%;  Pred. No. 2.7e-193;
  Matches  532;  Conservative  167;  Mismatches  357;  Indels   89;  Gaps   26;

Qy        161 LCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVC 220
```

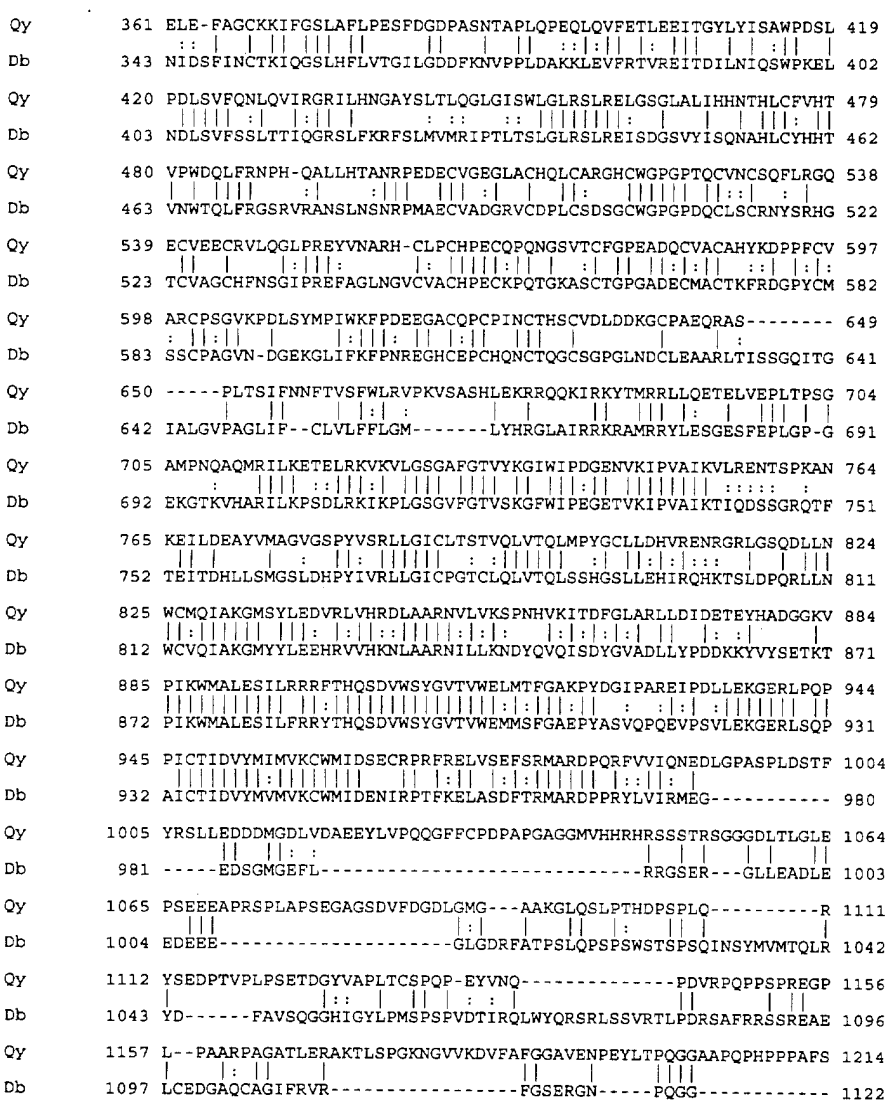


```
Db          3 LCFADTIHWQDIVRNPWASNFTLVPTNGSSGCGRCHKSCTG-RCWGPTENHCQTLTKTVC 61

Qy        221 AGGC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTF 279

Db         62 AEQCDGRCYGPYVSDCCHRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTF 121

Qy        280 ESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCA 339

Db        122 QLEHNHNAKYTYGAFCVKKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICP 179

Qy        340 RVCYGLGMQYIKANSKFIGITELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQ 398

Db        180 KACDGIGTGSL-VSAQTVDSSNIDKFINCTKINGNLIFLVTGIHGDPYHTIAAINPEKLN 238

Qy        399 VFETLEEITGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSL 458

Db        239 IFQTVREITGYLNIQSWPENMTDFRVFSNLVTIGGRALYSGLSLLILKQQGITSLQFQSL 298

Qy        459 RELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARG 518

Db        299 KQISAGNIYITDNSNLCYYHTVNWTSLFSTPSQKTVIHRNKKAENCTADGMVCNELCSSD 358

Qy        519 HCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCF 577

Db        359 GCWGPGPDQCLSCKRFIRGRTCIESCNLYDGEFREFANGSVCMECDPQCEKMEDNMITCY 418

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637

Db        419 GPGPDHCTKCFHFKDGPNCVEKCPDGLQGANSF--IFKYADEDRECHPCHPNCTQGC--- 473

Qy        638 DDKGCPAEQRA--SPLTSIFNNFTVSFWLRVPKVSAS--------------HLEKRRQQ 680

Db        474 --RG-PASHDCIYYPWT---RQSTLPQHARTPLIAAGVIGGLFIIVIMGLTFAVYVRRKS 527

Qy        681 KIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWI 740

Db        528 IKKKRALRRFL-ETELVEPLTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWV 586

Qy        741 PDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQL 800

Db        587 PEGETVKIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQL 646

Qy        801 MPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVK 860

Db        647 MPHGCLLDYVHEHKDNIGSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVK 706

Qy        861 ITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGA 920

Db        707 ITDFGLARLLEGDEKEYNADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGG 766

Qy        921 KPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMA 980

Db        767 KPYDGIPTREIPDLLEKGERLPQPPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMA 826

Qy        981 RDPQRFVVIQNED-LGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGA 1039

Db        827 RDPQRYLVIQGDDRMKLPSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTS 885

Qy       1040 GGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS--PLAP-SEGAGSDVFDGDLGMGAAK 1096

Db        886 RTRIDSNRNQFVYRDGGYAAEQGV-PMPYRAPGCIIPEAPVAQGATAEIFEDTCCNGTLR 944

Qy       1097 GLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQP 1149

Db        945 KQVATLAKEDSSTQRYSADPTVFIPERVIRGELDEDGYMTPMRDKPKTDYLNPVEENPFV 1004

Qy       1150 PSPREGPLPAA-RPAGATLERAKTLSPGKNGVVKDVF--------AFGGAVENPEYLTPQ 1200

Db       1005 SRRKNGDLQAVDNPEYHN-------APNGQPKAEDEYVNEPLYLNTFANTLENAEYL--- 1054

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGA--PPSTFKGTPT--------------AE 1244

Db       1055 -----KNNLPEKAKKAFDNPDYWNHSLPPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAE 1109

Qy       1245 NPEYL 1249

Db       1110 NPEYL 1114
```

RESULT 6
P79754
ID   P79754   PRELIMINARY;   PRT; 1328 AA.
AC   P79754;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ErbB3.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99177347; PubMed=10077531;
RA   Gellner K., Brenner S.;
RT   "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT   rubripes.";
RL   Genome Res. 9:251-258(1999).
DR   EMBL; AF056116; AAC34391.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   1328 AA; 148613 MW; A333039258B647E9 CRC64;

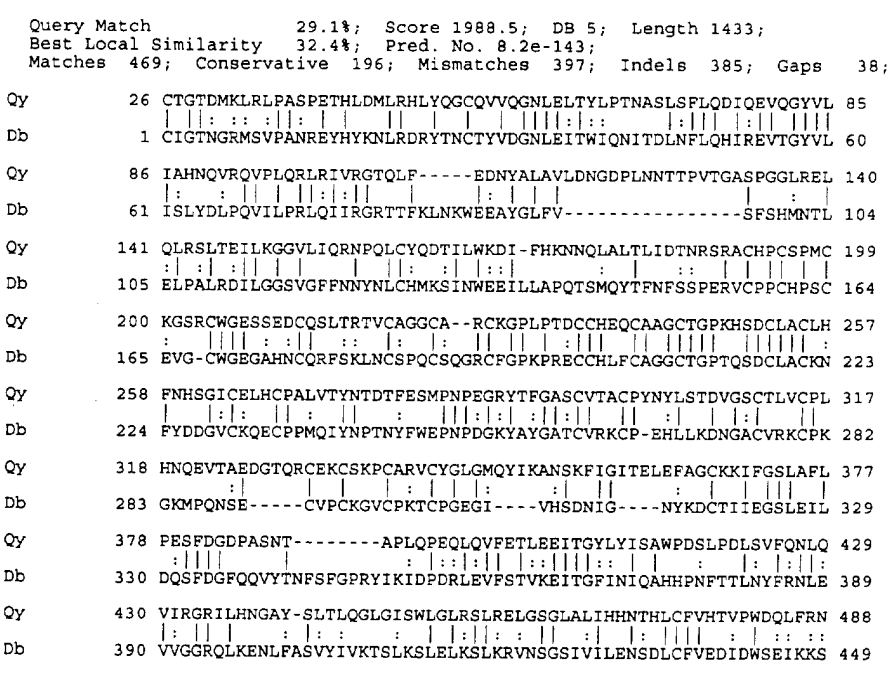
  Query Match   33.1%; Score 2260; DB 13; Length 1328;
  Best Local Similarity   39.9%; Pred. No. 1.2e-163;
  Matches 516; Conservative 155; Mismatches 418; Indels 204; Gaps   33;

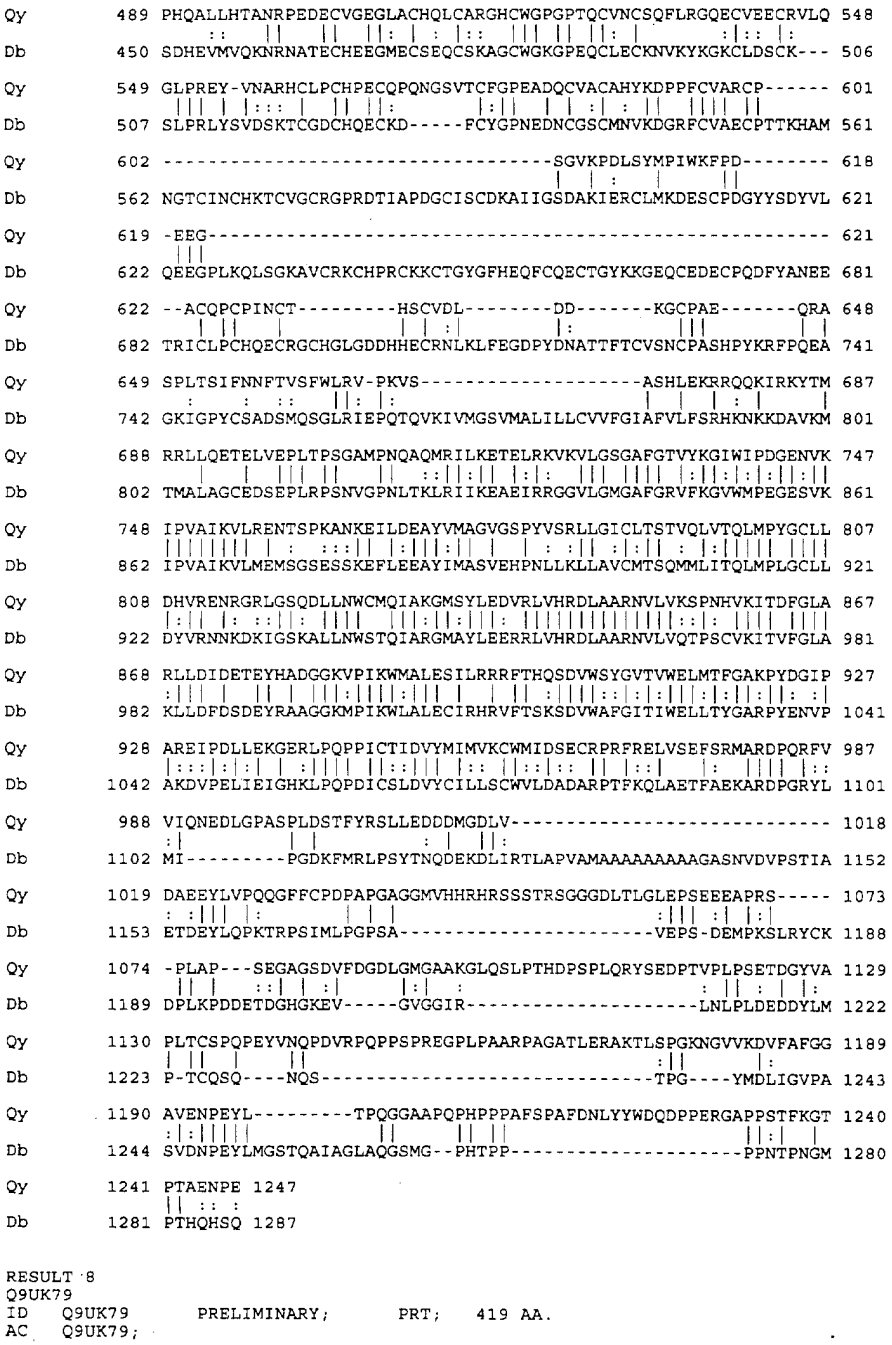
```
Qy          9 WGLLLALLPP--GAASTQ----VCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLEL 62

Db          4 WRLILMCVASRLRAASSQTQEAVCPGTQNGLSSTGSQENQYNLNKDRYKGCEIIMGNLEI 63

Qy         63 TYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDP 122

Db         64 TQIESNWDFSFLKTIREVTGYVLIAMNHFQEIPLGQLRVIRGNSLYERRFALSVFLN--- 120

Qy        123 LNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALT 182

Db        121 ----YPKDG--PSGLNQLGLMNLTEILDGGVQIINNKYLRYGPWVYWRDII-RNNDAPIE 173

Qy        183 LIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCA 241

Db        174 IQFNGERGVCH---KSC-GNYCWGPGKDQCQILTKTVCAPQCNDRCFGTSPRDCCHIECA 229

Qy        242 AGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY 301

Db        230 AGCKGPLDTDCFACRLFNDSGACVPQCPQTLIYNKQTFQMETNPNAKYQYGSICVSQCPT 289

Qy        302 NYLSTDVGSCTLVCPLHNQEVTAEDGTQR-CEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db        290 HFV-VDGSSCVSVCPPDKMEV--ERGSQRQCELCSGLCPKVCEGTGAE----QRQTVDSS 342
```



RESULT 7
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

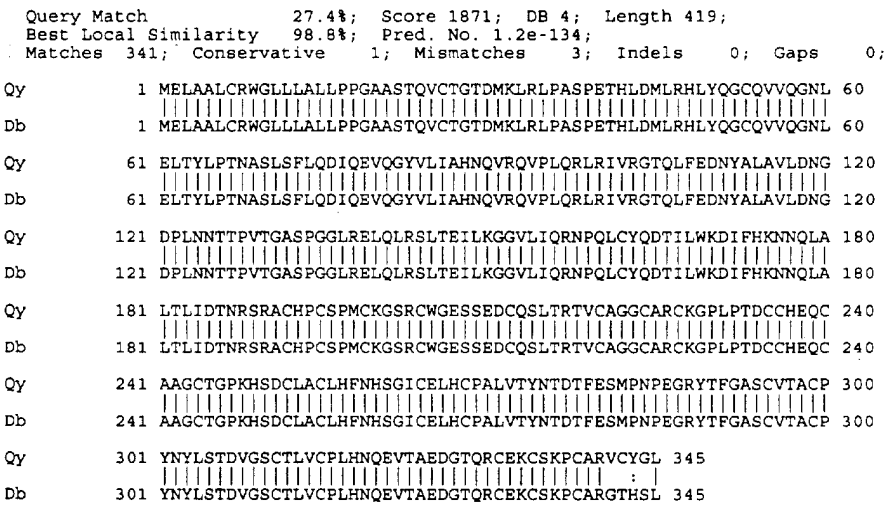
Query Match 29.1%; Score 1988.5; DB 5; Length 1433;
Best Local Similarity 32.4%; Pred. No. 8.2e-143;
Matches 469; Conservative 196; Mismatches 397; Indels 385; Gaps 38;





RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.4%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 1.2e-134;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;



RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.

```
DR   EMBL; BC027080; AAH27080.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   367 AA;  40163 MW;  0BE03395F9E101B0 CRC64;

  Query Match             25.4%;  Score 1739;  DB 11;  Length 367;
  Best Local Similarity   88.0%;  Pred. No. 1.3e-124;
  Matches  323;  Conservative   11;  Mismatches   33;  Indels    0;  Gaps    0;

Qy        889 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        949 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1008
              ||||||||||||||||||||||||||||||||||||||||||||:||:||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSL 120

Qy       1009 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1068
              |||||||:|||||||||||||||| |||| | |    | |||||| |||||:||||||||||
Db        121 LEDDDMGELVDAEEYLVPQQGFFSPDPALGTGSTAHRRHRSSSARSGGGELTLGLEPSEE 180

Qy       1069 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1128
              | |||||||||||||||||||| :|  ||||||  || ||||||||||:||| ||||||
Db        181 EPPRSPLAPSEGAGSDVFDGDLAVGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYV 240

Qy       1129 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1188
              ||| ||||||||||:|||| |   ||| |  |||||||| ||||||||||||||||||||
Db        241 APLACSPQPEYVNQPEVRPQSPLTPEGPPPPIRPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1189 GAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1248
              |||||||| |: | | |||| ||||||||||||||:  |:| |||||:||||||||||
Db        301 GAVENPEYLAPRAGTASQPHPSPAFSPAFDNLYYWDQNSSEQGPPPSTFEGTPTAENPEY 360

Qy       1249 LGLDVPV 1255
              |||||||
Db        361 LGLDVPV 367


RESULT 10
Q8WYV0
ID   Q8WYV0      PRELIMINARY;      PRT;   412 AA.
AC   Q8WYV0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 44.7 kDa protein.
GN   PP3659.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA   Wan D.F., Gu J.R.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF318349; AAL55856.1; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   412 AA;  44702 MW;  034397FF3F27D2BC CRC64;

  Query Match             24.8%;  Score 1697.5;  DB 4;  Length 412;
  Best Local Similarity   80.5%;  Pred. No. 2.3e-121;
  Matches  330;  Conservative    5;  Mismatches   22;  Indels   53;  Gaps    4;

Qy        889 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        949 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 120

Qy       1009 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 180

Qy       1069 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 240

Qy       1129 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFG 1188
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        241 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1189 GAVENPEYLTPQGGAAPQP----------------------------HPPPA---FSPAFDNL 1220
              |||||||||||||||| |                            || |:      :
Db        301 GAVENPEYLTPQGGAALSPTLLLPSAQPSTTSITGTRTHQSGGLHPAPSKGHLRQRTQST 360

Qy       1221 YYWD-QDPPER-----------------------GAPPSTFKGTPTAEN 1245
              : |  |  ||                       ||:| :||  | |
Db        361 WVWTCQCEPEGQVRRSPDVSSGSREGLTSAGIKRWEGPPTTSRGTCHARN 410


RESULT 11
Q86712
ID   Q86712      PRELIMINARY;      PRT;   729 AA.
AC   Q86712;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyprotein.
GN   POLYPROTEIN.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60725.1; -.
DR   HSSP; P03322; 1A6S.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR004028; Retro_M.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02813; Retro_M; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   729 AA;  80649 MW;  84D2F6914EFE1D63 CRC64;

  Query Match             24.5%;  Score 1675;  DB 15;  Length 729;
  Best Local Similarity   54.0%;  Pred. No. 2.8e-119;
  Matches  352;  Conservative   68;  Mismatches  132;  Indels  100;  Gaps   13;

Qy        569 PQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPI 628
              |:  :    ||  | |: |||: | | ||  ||:||  :   : :||: |    || |
```

```
Db        141 PEETATPKTGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHP 197

Qy        629 NCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMR 688
              ||| |     :|||     |   ||          |       |  ||:  :|| |:|
Db        198 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLR 254

Qy        689 RLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKI 748
              ||||| |||||||||| |||| :|||||| :|||||||||||||||:|||:|| |||
Db        255 RLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKI 314

Qy        749 PVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLD 808
              ||||| ||| |||||||||||||||| | :| | |||||||||||||:|||||||||||
Db        315 PVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCLLD 374

Qy        809 HVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLAR 868
              ::||::  :||| |||||:|||||:|||: |||||||||||||||:| ||||||||||:
Db        375 YIREHKDNIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAK 434

Qy        869 LLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPA 928
              ||  || ||||:|||||||||||||| | :|||||||||||||||||||:||||||||
Db        435 LLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPA 494

Qy        929 REIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVV 988
              ||  :|||||||||||||||||||||||:: ||:||||::|||:|||||  |::|
Db        495 SEISSVLEKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLV 554

Qy        989 IQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRH 1047
              || :| :   || || |||:|:|::|| |:|||:|||| ||||
Db        555 IQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF---------------- 598

Qy       1048 RSSSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLP 1102
              | ||                    |:||     | |  : ::  |        :  |  |
Db        599 NSPST-----------------SRTPLLSSLSATSNNSATNCID--------RNGQGHP 632

Qy       1103 THDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAA 1160
                : | :|||| |||     |:  ||::      | ||||||  : |: ||
Db        633 VREDSFVQRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS--------- 675

Qy       1161 RPAGATLERAKTLSPGKNGVVKDVF---------------AFGGAVENPEYL 1197
                      : :| :  ::              :   ||:|||||
Db        676 -----------TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 715


RESULT 12
Q86714
ID   Q86714      PRELIMINARY;      PRT;   567 AA.
AC   Q86714;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   V-erbB protein (Fragment).
GN   V-ERBB.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60727.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Tyrosine-protein kinase.
FT   NON_TER       1       1
SQ   SEQUENCE   567 AA;  63390 MW;  C6D9CBA7ADF725E1 CRC64;

  Query Match             24.5%;  Score 1673;  DB 15;  Length 567;
  Best Local Similarity   54.6%;  Pred. No. 2.7e-119;
  Matches  351;  Conservative   66;  Mismatches  126;  Indels  100;  Gaps   13;

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |: |||: | | ||  ||:||  :   : :||: |    || |  |||  |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELV 697
                :|||     |   ||           |       |  ||:  :|| |:|||||  |||
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELV 114

Qy        698 EPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRE 757
              |||||||  |||| :||||||| :|||||||||||||||:|||:|| |||||||  |||
Db        115 EPLTPSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELRE 174

Qy        758 NTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRL 817
              |||||||||||||||| | :| | |||||||||||||:||||||||||::||::  :
Db        175 ATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNI 234

Qy        818 GSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEY 877
              ||| |||||:|||||:|||: |||||||||||||||:| ||||||||||:||  || ||
Db        235 GSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEY 294

Qy        878 HADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 937
              ||:|||||||||||| | :|||||||||||||||||:|||||||| ||  :|||
Db        295 HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354

Qy        938 GERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGP 996
              ||||||||||||||||||||||:: ||:||||::|||:|||||  |::|||  :| :
Db        355 GERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHL 414

Qy        997 ASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGG 1056
              || || |||:|:|::|| |:|||:|||| ||||                  | ||
Db        415 PSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST---- 454

Qy       1057 GDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQR 1111
                           |:||     | |  : ::  |        :  |  |  : | :||
Db        455 --------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQR 492

Qy       1112 YSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLER 1169
              || |||     |:  ||::      | ||||||  : |: ||
Db        493 YSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS----------------- 526

Qy       1170 AKTLSPGKNGVVKDVF---------------AFGGAVENPEYL 1197
                    :| :  ::              :   ||:|||||
Db        527 ---TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 566


RESULT 13
Q64895
ID   Q64895      PRELIMINARY;      PRT;   962 AA.
AC   Q64895;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Gag,v-erb-A,v-erb-B protein.
GN   GAG,V-ERB-A,V-ERB-B.
OS   Avian erythroblastosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90206603; PubMed=1969616;
RA   Bruskin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT   "Six amino acids from the retroviral gene gag greatly enhance the
RT   transforming potential of the oncogene v-erb-B.";
```

```
RL   Oncogene 5:15-24(1990).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; X52209; CAA36459.1; -.
DR   EMBL; X52211; CAA36459.1; JOINED.
DR   HSSP; P10828; 2NLL.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW   Transcription regulation; Transferase; Tyrosine-protein kinase;
KW   Zinc-finger.
SQ   SEQUENCE   962 AA;  108320 MW;  3C5AED791E4E95CE CRC64;

  Query Match             23.5%;  Score 1608.5;  DB 15;  Length 962;
  Best Local Similarity   50.9%;  Pred. No. 5.2e-114;
  Matches  352;  Conservative   66;  Mismatches  156;  Indels  117;  Gaps   16;

Qy        541 VEECRVLQGLPRE-YVNAR-HCLP----------------------------CHPECQ 568
              :|:|:    |  | |:| | | :|                            |  |
Db        354 IEKCQESYLLAFEHYINYRKHNIPHFWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413

Qy        569 PQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPI 628
              ||       ||  | |: |||: | | ||  ||:||  :   : :||: |    || |
Db        414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHP 465

Qy        629 NCTHSCVDLDDKGCPAEQRASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMR 688
              |||  |     :|||     |   ||          |      |  ||:  :|| |:|
Db        466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLR 522

Qy        689 RLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKI 748
              ||||| |||||||||  |||| :|||||| :|||||| ||||||||:|||:|| | |
Db        523 RLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTI 582

Qy        749 PVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLD 808
              ||||| ||| |||||||||||||||| | :|:| |||||||||||||:||||||||||
Db        583 PVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLD 642

Qy        809 HVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLAR 868
              ::||::  :||| ||||:|||||:|||:  :|||||||||||:| ||||||||||:
Db        643 YIREHKDNIGSQYLLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAK 702

Qy        869 LLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPA 928
              |  || ||||:|||||||||||||| | :|||||||||||||||||:|||||||
Db        703 QLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPA 762

Qy        929 REIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVV 988
              ||  :|||||||||||||||||||||||||  :: ||:||||::|||:|||||  |::|
Db        763 SEISSVLEKGERLPQPPICTIDVYMIMVKCWMSGADSRPKFRELIAEFSKMARDPPRYLV 822

Qy        989 IQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRH 1047
              || :|  :   || || |||:|:|::|| |:|||:|||| ||||
Db        823 IQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF---------------- 866

Qy       1048 RSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPS 1107
              | ||                  |:||  |  | |:         | | :    |
Db        867 NSPST-----------------SRTPLLSSLSATSN-------NSATKCIDRNGGH--- 898

Qy       1108 PLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT- 1166
                          |  ||::     | ||||||  : |: ||              |
Db        899 -------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTA 937

Qy       1167 LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1197
                : :    |  :|       :   ||:|||||
Db        938 ISKLPMDSRYQN-------SHSTAVDNPEYL 961


RESULT 14
Q85468
ID   Q85468      PRELIMINARY;      PRT;   545 AA.
AC   Q85468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS   Avian erythroblastosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88217326; PubMed=2897102;
RA   Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT   "Common site of mutation in the erbB gene of avian erythroblastosis
RT   virus mutants that are temperature sensitive for transformation.";
RL   Oncogene Res. 1:265-278(1987).
DR   EMBL; X06943; CAA30024.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   545 AA;  60899 MW;  140DCE8CCA0F8AF4 CRC64;

  Query Match             23.4%;  Score 1600;  DB 15;  Length 545;
  Best Local Similarity   54.1%;  Pred. No. 9.9e-114;
  Matches  339;  Conservative   63;  Mismatches  135;  Indels   90;  Gaps   13;

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |: |||: | | ||  ||:||  :   : :||: |    || |  ||| |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIFNNFTVSFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELV 697
              :|||     |   ||           |      |  ||:  :|| |:|||| |||
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELV 114

Qy        698 EPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRE 757
              |||||||  |||| :||||||| :|||||| |||||||||:|||:|| | |||||| |||
Db        115 EPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELRE 174

Qy        758 NTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRL 817
              |||||||||||||||| | :|:| ||||||||||||:|||||||||||::||::  :
Db        175 ATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNI 234

Qy        818 GSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEY 877
              ||| |||||:||||||:|||:  |||||||||||||:|  |||||||||: |  || ||
Db        235 GSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQLGADEKEY 294

Qy        878 HADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 937
              ||:|||||||||||| | :|||||||||||||||||:|||||||| ||   :|||
Db        295 HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354

Qy        938 GERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGP 996
              |||||||||||||||||||| |:: ||:||||::|||:|||| |::||| :| :
Db        355 GERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHL 414
```

```
Qy        997 ASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGG 1056
              || || |||:|:|::|| |:|||:||||| ||||                  | ||
Db        415 PSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST---- 454

Qy       1057 GDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQR 1111
                        |:||     | |  : ::  | : |             |
Db        455 --------------SRTPLLSSLSATSNNSATNCIDRNGG-----------H------- 481

Qy       1112 YSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERA 1170
                    |   ||::     | ||||||  : |: ||    :     | : :
Db        482 ---------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQIQIYNYISLTAISKL 524

Qy       1171 KTLSPGKNGVVKDVFAFGGAVENPEYL 1197
                  |  :|       :   ||:|||||
Db        525 PMDSRYQN-------SHSTAVDNPEYL 544


RESULT 15
Q9WVF5
ID   Q9WVF5      PRELIMINARY;      PRT;   655 AA.
AC   Q9WVF5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor (Epidermal growth factor receptor
DE   isoform 3).
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT   Carboxy-Terminal Truncated Receptors.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative Egfr transcripts encoding truncated receptor
RT   isoforms.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AF124513; AAD44149.1; -.
DR   EMBL; AF275366; AAG28047.1; -.
DR   EMBL; AF275364; AAG28047.1; JOINED.
DR   EMBL; AF275365; AAG28047.1; JOINED.
DR   EMBL; AK004944; BAB23688.1; -.
DR   EMBL; AK004883; BAB23641.1; -.
DR   EMBL; AK004911; BAB23662.1; -.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR002174; Furin-like.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 3.
KW   Receptor.
SQ   SEQUENCE   655 AA;  72906 MW;  6B34063B1BC928CB CRC64;

  Query Match             22.0%;  Score 1506.5;  DB 11;  Length 655;
  Best Local Similarity   44.5%;  Pred. No. 1.8e-106;
  Matches  285;  Conservative   97;  Mismatches  233;  Indels   25;  Gaps    9;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
              || ||    | |     :|| ||  :|    : | |   |: :|  |:|| ||||:||:  |
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
                ||||: |||| |||||| | | ::||: |:|:||  |:|: ||||:| |
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
               | :  ||||| :|:| ||| | |    || ||  ||| |:||    |  |  : |
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
                :  :|  | | |     |||   |:|| ||:':||  |: ||:|  |:|||| |||||
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              ||||: |||| | |      |:  || |: ||  |::   ||||:|:|||:||  || ||
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              : || |||   |     ||  ||| ::|:||  || :|| |:|:   | ::  |  | ::
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
               |  | | | | || :| ||  : | || | :|:: :|::||||:| | ||||:  ||
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
              |:||::|||   :| :|| : || |: |||||:|:  |  :|  | :||: :|: |
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
               :||  |:|       || | :|       |: ||:   |||| |  ||:|    ||:||||
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              :| :|:| |||:|    |: ||||| ||  ::|| |    | |: |||| | | ||  ||:
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC 642
              |:  : : : :||: |    |  |||: |     :||
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGC 636


Search completed: July 22, 2003, 09:24:19
Job time : 51.8246 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.9339 Seconds
(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-632-652-14
Perfect score: 6834
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6627	97.0	1255	21	AAY92620	Human heregulin 2
2	6627	97.0	1255	22	AAE12130	Human tyrosine kin
3	6627	97.0	1255	22	AAB60167	HER2 transgene pla
4	6627	97.0	1255	23	AAU74545	Human HER2 (ErbB2)
5	6623	96.9	1255	17	AAW01111	HER-2/neu protein.
6	6623	96.9	1255	20	AAW92406	Human HER-2/neu on
7	6623	96.9	1255	21	AAB21198	Human HER-2/neu pr
8	6623	96.9	1255	21	AAY84780	Amino acid sequenc
9	6623	96.9	1255	22	AAB85458	Human HER-2/neu pr
10	6623	96.9	1255	22	AAG88267	HER2/neu amino aci
11	6623	96.9	1255	23	AAE24067	Human Her-2 protei
12	6623	96.9	1255	23	AAE20479	Human Her-2/neu pr
13	6623	96.9	1255	23	AAM51143	Human Her-2/neu on
14	6623	96.9	1255	23	AAU77114	Human Her-2/neu po
15	6580	96.3	1433	14	AAR39568	Sequence of c-erbB
16	6457	94.5	1223	23	AAU98923	Human breast cance
17	6306	92.3	1200	21	AAB21208	Human HER-2/neu pr
18	5869.5	85.9	1256	21	AAB21199	Rat HER-2/neu prot
19	5869.5	85.9	1256	23	AAM51144	Rat Her-2/neu onco
20	5842.5	85.5	1256	21	AAB21206	Mouse Her-2/neu pr
21	5842.5	85.5	1256	22	AAG62860	Amino acid sequenc
22	5842.5	85.5	1256	23	AAM51151	Mouse Her-2/neu on
23	4814	70.4	919	21	AAB21203	Human HER-2/neu fu
24	4814	70.4	919	23	AAM51148	Her-2/neu extracel
25	4064.5	59.5	920	23	AAM51152	Mouse Her-2/neu ex
26	4064.5	59.5	926	23	AAM51153	Mouse Her-2/neu ex
27	3698	54.1	712	21	AAB21204	Human HER-2/neu fu
28	3698	54.1	712	23	AAM51149	Her-2/neu extracel
29	3556	52.0	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	51.9	653	21	AAB21200	Extracellular HER-
31	3550	51.9	653	23	AAM51145	Human Her-2/neu on
32	3512	51.4	645	22	AAB60408	Human ErbB2 oncopr
33	3512	51.4	645	22	AAB61593	Human ErbB2 extrac
34	3447	50.4	951	21	AAY44993	DC8scFv-erbB2EC fu
35	3344	48.9	624	11	AAR08222	Extracellular port
36	3113	45.6	1210	21	AAB19259	Amino acid sequenc
37	3113	45.6	1210	21	AAY50616	Human EGF receptor
38	3113	45.6	1210	23	AAE23019	Human Her-1 protei
39	3113	45.6	1210	23	AAM50768	Human epidermal gr
40	3111	45.5	1210	22	AAB68420	Amino acid sequenc
41	3088	45.2	589	23	AAE20484	Human protein for
42	3084	45.1	600	23	AAE20482	Human protein for
43	3083	45.1	583	23	AAE20483	Human protein for
44	3083	45.1	587	23	AAE20481	Human protein for
45	3072	45.0	1210	23	ABP51768	Human epidermal gr

ALIGNMENTS

RESULT 1
AAY92620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

```
FT                   /label= Cysteine_rich_domain
FT   Region          210..224
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          250..264
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          324..483
FT                   /label= Ligand_binding_domain
FT   Region          325..339
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          369..383
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          465..479
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          484..623
FT                   /label= Cysteine_rich_domain
FT   Region          579..593
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          624..654
FT                   /label= Transmembrane_domain
FT   Region          632..652
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          653..667
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          655..1010
FT                   /label= Tyrosine_kinase_domain
FT   Region          661..675
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          695..709
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          710..730
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          1011..1235
FT                   /label= C-terminal_domain
XX
PN   WO200020027-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;    99WO-DK00525.
XX
PR   05-OCT-1998;    98DK-0001261.
PR   20-OCT-1998;    98US-0105011.
XX
PA   (MEBI-) M & E BIOTECH AS.
XX
PI   Steinaa L,  Mouritsen S,  Nielsen KG,  Haaning J,  Leach D,  Dalum I;
PI   Gautam A,  Birk P,  Karlsson G;
XX
DR   WPI; 2000-349917/30.
DR   N-PSDB; AAA09455.
XX
PT   Inducing immune responses to weakly immunogenic, tumor associated
PT   peptide antigens for the treatment of breast and prostate cancer
XX
PS   Claim 62; Page 193-198; 220pp; English.
XX
CC   This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC   Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC   response.  Subdominant CTL epitopes, antibody binding regions and
CC   cysteine residues involved in disulfide bonds are preserved in the
CC   immunogenized forms.  Regions suitable for the insertion of foreign T
CC   helper epitopes were identified (see features table).  The method
CC   is used for inducing immune responses against weakly immunogenic
CC   cell-associated peptide antigens (PA) such as those associated with
CC   cancers (self-proteins), e.g. human prostate specific membrane antigen
CC   (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC   The method comprises effecting simultaneous presentation by antigen
CC   producing cells (APCs) of the animals immune system of: (1) at least 1
CC   CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC   B-cell group derived from the cell-associated PA; and (2) at least 1
CC   first T helper cell group which is foreign to the animal.  Analogues of
CC   human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC   part of all known and predicted CTL and B-cell epitopes of the respective
CC   PA and including at least one foreign T helper epitope are also claimed.
CC   The method is used to treat prostate, prostate/breast or breast cancer
CC   when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ   Sequence   1255 AA;

  Query Match             97.0%;  Score 6627;  DB 21;  Length 1255;
  Best Local Similarity   97.4%;  Pred. No. 0;
  Matches 1222;  Conservative    6;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                : |  |    | ||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
```

```
Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 2
AAE12130
ID   AAE12130 standard; Protein; 1255 AA.
XX
AC   AAE12130;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human tyrosine kinase-type receptor, HER-2.
XX
KW   Therapeutic compound; major histocompatibility complex; vaccine;
KW   antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW   adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW   antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          774..782
FT                   /note= "Antigenic epitope"
XX
PN   WO200168677-A2.
XX
PD   20-SEP-2001.
XX
PF   16-MAR-2001; 2001WO-US40328.
XX
PR   16-MAR-2000; 2000US-0527487.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Nicolette CA;
XX
DR   WPI; 2001-616284/71.
DR   N-PSDB; AAD19731.
XX
PT   Novel synthetic therapeutic compound for inducing immune response and
PT   for use in adoptive immunotherapy, has enhanced binding to major
PT   histocompatibility molecules and enhanced immunoregulatory properties
PT   -
XX
PS   Claim 4; Page 63-67; 69pp; English.
XX
CC   The invention relates to synthetic therapeutic compounds (antigenic
CC   peptides) with enhanced binding to major histocompatibility complex
CC   (MHC) molecules and enhanced immunoregulatory properties relative
CC   to their natural counterparts. Compounds of the invention are useful
CC   for inducing an immune response in a subject and for use in adoptive
CC   immunotherapy. They are useful as components of anti-cancer vaccines
CC   and to expand immune effector cells that are specific for cancers
CC   characterised by expression of the breast cancer antigen, HER-2.
CC   Polynucleotides that encode peptides of the invention are useful as
CC   hybridisation probes and as primers for the detection of genes of gene
CC   transcripts that are expressed in antigen presenting cells (APCs), to
CC   confirm transduction of polynucleotides into host cells. The present
CC   sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC   of the invention are designed based on the HER-2 antigenic peptide
CC   (774-782).
XX
SQ   Sequence   1255 AA;
```

```
  Query Match           97.0%;  Score 6627;  DB 22;  Length 1255;
  Best Local Similarity 97.4%;  Pred. No. 0;
  Matches 1222;  Conservative    6;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
```

```
Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 3
AAB60167
ID   AAB60167 standard; Protein; 1255 AA.
XX
AC   AAB60167;
XX
DT   03-APR-2001  (first entry)
XX
DE   HER2 transgene plasmid construct encoded protein.
XX
KW   Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW   antibody.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200100244-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-US17229.
XX
PR   25-JUN-1999;   99US-0141316.
PR   16-MAR-2000; 2000US-0189844.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Erickson S,  Schwall R;
XX
DR   WPI; 2001-061962/07.
DR   N-PSDB; AAF24297.
XX
PT   Treating tumors, particularly breast cancers, which overexpress an ErbB
PT   receptor and does not respond to an anti-ErbB antibody, comprises
PT   conjugating the antibody to a maytansinoid -
XX
```

```
PS   Example 3; Fig 4; 92pp; English.
XX
CC   The present invention provides a method of treating cancer by
CC   administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC   particular, the antibody is directed against ErbB2 (also known as HER2
CC   and p185neu). The method is particularly useful in the treatment of
CC   breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC   colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ   Sequence   1255 AA;
```

Query Match 97.0%; Score 6627; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              : |  |    | ||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoelic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
PI Erickson S, Schwall R, Sliwkowski M;
XX
DR WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoelic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;

Query Match 97.0%; Score 6627; DB 23; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

```
Qy           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy         361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy         601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db         601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy         661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                 : |  |    | ||||||||||||||||||||||||||||||||||||||||||||
Db         661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy         721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy         781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
```

```
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 5
AAW01111
ID   AAW01111 standard; Protein; 1255 AA.
XX
AC   AAW01111;
XX
DT   01-JAN-1997  (first entry)
XX
DE   HER-2/neu protein.
XX
KW   HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW   breast cancer; ovary cancer; colon cancer; lung cancer;
KW   prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          676..1255
FT                   /label= Intracellular_domain
FT                   /note= "claimed domain, useful for immunisation"
XX
PN   WO9630514-A1.
XX
PD   03-OCT-1996.
XX
PF   28-MAR-1996;   96WO-US01689.
XX
PR   31-MAR-1995;   95US-0414417.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Disis ML;
XX
DR   WPI; 1996-455361/45.
DR   N-PSDB; AAT40739.
XX
PT   DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT   treatment of malignancies with which the HER-2/neu oncogene is
PT   associated
XX
PS   Claim 2; Page 56-61; 71pp; English.
XX
CC   Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC   the product of the HER-2/neo oncogene (see also AAT40739).  The
CC   protein is over-expressed in various cancers, including breast,
CC   ovarian, colon, lung and prostate.  The intracellular domain of the
CC   protein can be used to immunise an animal against a malignancy with
CC   which the oncogene is associated.  The polypeptide can be produced
CC   in transformed host cells for use in immunisation.  Alternatively,
CC   animal cells are transfected in vivo or ex vivo with a viral vector
CC   that directs expression of the polypeptide.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6623;  DB 17;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              : |  |    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
```

```
Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX
AC AAW92406;
XX
DT 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 676..1255
FT /note= "region which elicits immune response"
XX
PN US5869445-A.
XX
PD 09-FEB-1999.
XX
PF 01-APR-1996; 96US-0625101.
XX
PR 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
PS Claim 3; Column 31-38; 26pp; English.
XX
CC This sequence represents the human HER-2/neu oncogene protein. A fragment
CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 20; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              : | | | |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
```

```
Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN WO200044899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX
PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;

```
  Query Match             96.9%;  Score 6623;  DB 21;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                 :  |   |    | ||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
```

```
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 8
AAY84780
ID   AAY84780 standard; Protein; 1255 AA.
XX
AC   AAY84780;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW   SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW   tumor cell proliferation; tissue degeneration; arthropathy;
KW   bone resorption; inflammatory disease; degenerative disorder;
KW   wound healing.
XX
OS   Homo sapiens.
XX
PN   WO200020579-A1.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-CA00912.
XX
PR   02-OCT-1998;   98US-0165192.
XX
PA   (UYMC-) UNIV MCMASTER.
XX
PI   Muller WJ,  Siegel PM;
XX
DR   WPI; 2000-303768/26.
DR   N-PSDB; AAA14812.
XX
PT   Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT   erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS   Claim 3; Fig 2; 60pp; English.
XX
CC   The present sequence represents a SPLICE erbB-2 receptor protein. The
CC   protein has an in-frame deletion of 16 amino acids, 2 of which are
CC   conserved cysteine residues, compared to the unspliced protein. The
CC   erbB-2 polynucleotide is used to construct probes for detecting
CC   disorders of cell transformation such as cancer. Antibodies to the
CC   protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC   (e.g. antisense oligonucleotides) which inhibit the expression of
CC   SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC   treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC   for treating conditions involving damaged cells including conditions
CC   in which degeneration of tissue occurs, such as arthropathy, bone
CC   resorption, inflammatory diseases, degenerative disorders of the
CC   central nervous system and wound healing.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match            96.9%;  Score 6623;  DB 21;  Length 1255;
  Best Local Similarity  97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
                 |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 9
AAB85458
ID   AAB85458 standard; Protein; 1255 AA.
XX
AC   AAB85458;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human HER-2/neu protein.
XX
KW   Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW   oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS   Homo sapiens.
XX
PN   WO200153463-A2.
XX
PD   26-JUL-2001.
XX
PF   19-JAN-2001; 2001WO-US01850.
XX
PR   21-JAN-2000; 2000US-0177545.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Cheever MA,  Hand-Zimmermann S;
XX
DR   WPI; 2001-476112/51.
DR   N-PSDB; AAH23392.
XX
PT   New antigen-presenting cells, useful as vaccines for eliciting or
PT   enhancing an immune response to HER-2/neu protein, particularly useful
PT   for treating or preventing cancer, e.g. breast cancer  -
XX
PS   Claim 2; Page 41-46; 49pp; English.
XX
CC   The invention provides an isolated antigen-presenting cell, which
CC   expresses at least an immunogenic portion of a polypeptide that produces
CC   an immune response to HER-2/neu protein. The antigen-presenting cells are
CC   useful as vaccines for eliciting or enhancing an immune response to
CC   HER-2/neu protein, particularly in treating or preventing malignancies in
CC   which the HER-2/neu oncogene is associated. Specifically, these are
CC   useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC   colon, lung or prostate cancers. The present sequence represents
CC   the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6623;  DB 22;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                 : |  |    | ||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
                 |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
```

```
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 10
AAG88267
ID   AAG88267 standard; Protein; 1255 AA.
XX
AC   AAG88267;
XX
DT   11-SEP-2001  (first entry)
XX
DE   HER2/neu amino acid sequence.
XX
KW   Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW   immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW   tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS   Homo sapiens.
XX
PN   WO200141787-A1.
XX
PD   14-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US33591.
XX
PR   10-DEC-1999;   99US-0458299.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
PI   Keogh E;
XX
DR   WPI; 2001-374995/39.
XX
PT   An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT   cellular immune responses for the prevention and treatment of cancer -
XX
PS   Disclosure; Page 15; 199pp; English.
XX
CC   The present invention describes isolated prepared HER2/neu epitopes (I).
CC   Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC   culture in vitro and binds to a complex of an epitope (I), bound to a
CC   human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC   and a second epitope and the peptide is less than 50 contiguous amino
CC   acids that have 100% identity with a native peptide sequence of HER2/neu;
CC   (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC   excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC   (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC   and immunostimulant activities, and can be used in vaccines. (I), (II)
CC   and (III) are useful for inducing cellular immune responses for the
CC   prevention and treatment of cancer. (I) and (II) are useful for
CC   monitoring or evaluating an immune response to a tumour-associated
CC   antigen when incubated with a T lymphocyte sample form a patient and
CC   detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC   based vaccines mean that immunosuppressive epitopes that may be present
CC   in whole antigens may be avoided. Selected epitopes may be combined to
CC   enhance immunogenicity. The possible pathological side effects caused by
CC   infectious agents or whole protein antigen is eliminated. The vaccine
CC   provides the ability to direct and focus an immune response to multiple
CC   selected antigens from the same pathogen. Epitope-based anti-tumour
CC   vaccines provides the opportunity to combine epitopes derived from
CC   multiple tumour-associated molecules addressing the problem of tumour-
CC   tumour variability and reducing the likelihood of tumour escape due to
CC   antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6623;  DB 22;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
                : |  |    | ||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 11
AAE24067
ID   AAE24067 standard; Protein; 1255 AA.
XX
AC   AAE24067;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Human Her-2 protein.
XX
KW   Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW   hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW   tumour; gene therapy; phosphorothioate backbone.
XX
OS   Homo sapiens.
XX
PN   WO200222636-A1.
XX
PD   21-MAR-2002.
XX
PF   12-SEP-2001; 2001WO-US28572.
XX
PR   15-SEP-2000; 2000US-0663834.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Cowsert LM;
XX
DR   WPI; 2002-471192/50.
DR   N-PSDB; AAD38904.
XX
PT   Novel antisense oligonucleotide which modulates the expression of Human
PT   Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT   inflammation or to prevent infection in humans -
XX
PS   Example 13; Page 95-107; 116pp; English.
XX
CC   The invention relates to antisense compounds targetted to a nucleic
CC   acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC   that specifically hybridises with and inhibits the expression of Her2.
CC   Antisense compounds of the invention are used for treating diseases or
CC   conditions associated with Her2 such as hyperproliferative disorders
CC   e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC   neural or cardiac cancer. They are also useful prophylactically e.g.
CC   to prevent or delay infection, inflammation and tumour formation. The
CC   invention is also used in gene therapy. The present sequence is human
CC   Her-2 protein.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6623;  DB 23;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              :  |  |    | |||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
```

```
                 |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 12
AAE20479
ID   AAE20479 standard; Protein; 1255 AA.
XX
AC   AAE20479;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Human Her-2/neu protein.
XX
KW   Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW   human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1021..1030
FT                   /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN   WO200214503-A2.
XX
PD   21-FEB-2002.
XX
PF   14-AUG-2001; 2001WO-US41733.
XX
PR   14-AUG-2000; 2000US-225152P.
PR   28-SEP-2000; 2000US-236428P.
PR   21-FEB-2001; 2001US-270520P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Hand-zimmermann S,  Cheever MA,  Foy TM,  Lodes MJ,  Kalos MD;
PI   Mcneill PD,  Vedvick TS;
XX
DR   WPI; 2002-280758/32.
DR   N-PSDB; AAD32743.
XX
PT   Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT   prevention and diagnosis of cancer, preferably breast cancer  -
XX
PS   Disclosure; Page 114-117; 129pp; English.
XX
CC   The invention relates to an isolated Her-2/Neu polypeptide composition
CC   effective for eliciting an immune response. The invention is useful for
CC   eliciting an immune response in a patient, where the patient is human
CC   leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC   The composition is useful for the therapy and diagnosis of cancer,
CC   preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC   and other compositions for the diagnosis, prevention and treatment of
CC   human malignancies, for stimulating and/or expanding T cells specific for
CC   Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC   patient. The invention is useful for stimulating a T cell response in a
CC   human patient, as probe or primer for nucleic acid hybridisation, to
CC   selectively form duplex molecules with complementary stretches of the
CC   entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC   length gene from a suitable library, and to direct expression of a
CC   polypeptide in appropriate host cells. The composition is useful in
CC   prophylactic or therapeutic applications and for the treatment of cancer,
CC   preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC   associated malignancies. The invention is useful in gene therapy. The
CC   present sequence is human Her-2/neu protein.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6623;  DB 23;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
D         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              : |  |    | ||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 13
AAM51143
ID   AAM51143 standard; Protein; 1255 AA.
XX
AC   AAM51143;
XX
DT   17-JUN-2002  (first entry)
XX
DE   Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX
KW   Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW   tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..653
FT                   /note= "extracellular domain"
FT   Domain          676..1255
FT                   /note= "intracellular domain"
FT   Domain          990..1255
FT                   /note= "phosphorylation domain"
XX
PN   WO200212341-A2.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-US24283.
XX
PR   03-AUG-2000; 2000US-0632507.
XX
PA   (CORI-) CORIXA CORP.
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Cheever MA,  Gheysen D;
XX
DR   WPI; 2002-241743/29.
DR   N-PSDB; ABA92250.
XX
PT   Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT   or enhancing an immune response to the protein, has Her-2/neu
PT   extracellular domain fused to Her-2/neu intracellular or
PT   phosphorylation domain  -
XX
PS   Claim 68; Fig 7; 141pp; English.
XX
CC   The present sequence is that of human Her-2/neu (p185 glycoprotein
CC   or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC   vaccines.  The Her-2/neu gene is amplified and p185 is overexpressed
CC   in a variety of cancers, including breast, ovarian, colon, lung and
CC   prostate cancer.  Her-2/neu is a member of the tyrosine kinase
CC   family of receptor-like glycoproteins.  It comprises an extracellular
CC   domain with homology to the epidermal growth factor receptor
CC   (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC   intracellular domain that also shows homology to EGFR.  Its
CC   overexpression correlates with a poor prognosis in breast and
CC   ovarian cancers.  The invention provides Her-2/neu fusion
CC   proteins, nucleic acids encoding them, viral vectors, and vaccines
CC   comprising the fusion proteins or nucleic acid molecules.  In
CC   preferred fusion proteins, the extracellular domain of a Her-2/neu
CC   protein is fused to a Her-2/neu intracellular domain or
CC   phosphorylation domain (or its DeltaPD fragment).  An immune
CC   response to Her-2/neu protein is elicited or enhanced by
CC   administering the fusion protein in the form of a vaccine, or by
CC   transfecting cells of an animal ex vivo with a nucleic acid
CC   encoding the fusion protein, and delivering the transfected cells
CC   to the animal.  The fusion proteins, nucleic acids, and isolated
CC   specific T-cells are useful for inhibiting the development of a
CC   cancer, especially breast, ovarian, colon, lung or prostate cancer
CC   in a patient.  T cells that specifically react with a Her-2/neu
CC   fusion protein can be used to remove tumour cells from a sample in
CC   order to inhibit the development of cancer in a patient.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match             96.9%;  Score 6623;  DB 23;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
               : |  |    | ||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
```

```
Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 14
AAU77114
ID   AAU77114 standard; Protein; 1255 AA.
XX
AC   AAU77114;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Human Her-2/neu polypeptide.
XX
KW   Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW   acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW   chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW   Hodgkin's lymphoma; T cell therapy.
XX
OS   Homo sapiens.
XX
PN   WO200213847-A2.
XX
PD   21-FEB-2002.
XX
PF   13-AUG-2001; 2001WO-US25408.
XX
PR   14-AUG-2000; 2000US-0638280.
PR   28-SEP-2000; 2000US-0675904.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Gaiger A,  Cheever MA,  Hand-zimmermann S;
XX
DR   WPI; 2002-280741/32.
DR   N-PSDB; ABK10730.
XX
PT   Inhibiting haematological malignancy development by administering
PT   polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT   encoding the polypeptide, or antigen presenting cells expressing the
PT   polypeptide -
XX
PS   Disclosure; Page 71-74; 74pp; English.
XX
CC   The invention relates to a method for inhibiting development of
CC   haematological malignancy in a patient by administering a polypeptide
CC   comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC   encoding the polypeptide. Antigen presenting cells that express the
CC   protein can also be administered. The sequences are used for inhibiting
CC   development of haematological malignancy such as acute myelogenous
CC   leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC   leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC   lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match           96.9%;  Score 6623;  DB 23;  Length 1255;
  Best Local Similarity 97.3%;  Pred. No. 0;
  Matches 1221;  Conservative    6;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              : |  |    | ||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900
```

```
Db        841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Db        901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Db        961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Db       1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells. such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isolelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 96.3%; Score 6580; DB 14; Length 1433;
Best Local Similarity 96.7%; Pred. No. 0;

Matches 1214; Conservative 8; Mismatches 33; Indels 0; Gaps 0;

```
Qy           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQVCQVVQGNL 60

Qy          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEHCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy         361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db         361 IQEFAGCKRIFGSLAFLPESFDGDPASNTAPLQPEHLQVFETLEQITGYLYISAWPDSLP 420

Qy         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLSFVHTV 480

Qy         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db         541 VEECRVLQGLPREYVNASHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy         601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIFNNFTV 660

Db         601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy         661 SFWLRVPKVSASHLEKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db         661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy         721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Db         721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSP 780

Qy         781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Db         781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVR 840

Qy         841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Db         841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFT 900

Qy         901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Db         901 HQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWM 960

Qy         961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Db         961 IDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020

Qy        1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080

Db        1021 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1080
```

```
Qy       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy       1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1141 NQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQ 1200

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

Search completed: July 22, 2003, 09:16:28
Job time : 41.9339 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.25 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-653-675-14
Perfect score: 6814
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6604	96.9	1255	1	A24571	protein-tyrosine k
2	5811	85.3	1260	1	TVRTNU	protein-tyrosine k
3	5801.5	85.1	1254	2	I48161	p-185 precursor -
4	3069	45.0	1210	1	GQHUE	epidermal growth f
5	3040	44.6	1210	2	A53183	epidermal growth f
6	3017.5	44.3	1223	1	TVCHLV	epidermal growth f
7	2908.5	42.7	1308	2	A47253	epidermal growth f
8	2598	38.1	1166	1	S06142	protein-tyrosine k
9	2380.5	34.9	1342	2	A36223	kinase-related tra
10	2292.5	33.6	1339	2	JC4387	epidermal growth f
11	1689.5	24.8	698	1	TVFVLV	protein-tyrosine k
12	1626	23.9	604	1	TVYUH	protein-tyrosine k
13	1599.5	23.5	1330	1	GQFFE	epidermal growth f
14	1570	23.0	544	2	S35745	protein-tyrosine k
15	1563	22.9	545	2	S00727	kinase-related tra
16	1546	22.7	540	2	B44776	protein-tyrosine k
17	1544	22.7	540	1	TVFVEB	protein-tyrosine k
18	1509	22.1	644	2	A36325	epidermal growth f
19	1287	18.9	1323	2	E88257	protein let-23 [im
20	1287	18.9	1374	2	S70712	protein-tyrosine k
21	1194	17.5	1369	2	S70713	protein-tyrosine k
22	1142	16.8	1717	1	A45558	epidermal growth f
23	1126	16.5	527	2	A42032	epidermal growth f
24	977.5	14.3	843	2	A27131	epidermal growth f
25	806.5	11.8	346	2	S13807	protein-tyrosine k
26	754.5	11.1	311	2	S13808	protein-tyrosine k
27	724	10.6	1363	2	T43220	insulin-like growt
28	706	10.4	1383	2	A36080	insulin receptor p
29	705.5	10.4	1372	2	A34157	insulin receptor p
30	705	10.3	1382	1	INHUR	insulin receptor p
31	691.5	10.1	1300	2	A36502	insulin receptor-r
32	685	10.1	1477	2	T18534	protein-tyrosine k
33	680	10.0	1607	2	T43212	insulin-like growt
34	667	9.8	1268	2	B36502	insulin receptor-r
35	638	9.4	1367	1	IGHUR1	insulin-like growt
36	631.5	9.3	2148	1	A56081	insulin receptor -
37	625	9.2	2101	2	S57245	insulin receptor (
38	624	9.2	1371	2	A33837	insulin-like growt
39	622	9.1	1390	2	T30346	insulin receptor -
40	604	8.9	1114	1	S05582	protein-tyrosine k
41	592	8.7	987	2	A54092	protein-tyrosine k
42	584.5	8.6	952	2	I50612	protein-tyrosine k
43	584.5	8.6	1091	2	S33596	protein-tyrosine k
44	576.5	8.5	984	2	A39753	protein-tyrosine k
45	572.5	8.4	801	4	TVHURE	transforming prote

ALIGNMENTS

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming proteir
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito,
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal gro
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/e
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeb
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares c
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517,'RALL',522,'S',524-654,'V',656-1169,'A',1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:g183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcrip
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.3%; Score 5811; DB 1; Length 1260;
Best Local Similarity 85.2%; Pred. No. 5.1e-232;
Matches 1074; Conservative 57; Mismatches 118; Indels 12; Gaps 4;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||||
Db          4 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 63

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| |||||||||||||||:|||||::|||||||||||||||| ||||||||
Db         64 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 123

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||: |||||||| :||||:| |||||
Db        124 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 183

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |   |||||||| ||:| || : |||| ||| || |:|  ||||||| |||||||||
Db        184 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 243

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |
Db        244 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 303

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||||||||||||||:::::
Db        304 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 363

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
                 || |||||||||||||||||:|  |||:|||||||||||||||||||||||||||||
Db        364 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 423

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              ||||||||::|||||||:|||||||||  |||||||||||||||| | |||||||
Db        424 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 483

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||: ||||:: ||  || |: ||| |||||||||||||| |||||
Db        484 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 543

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : |||||||||||| | |||| |||| |||||||   |||
Db        544 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 603

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAV 658
              |||||||||||||||||:|||||  |||||||||||||||::|||||||||:|  |:: |
Db        604 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRASPVTFIIATV 663

Qy        659 VGILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRI 712
              |:|| ::| || ||||         :|:   |:  ::    :|||||||||||||||||
Db        664 EGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRI 719

Qy        713 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 772
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 779

Qy        773 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMS 832
              |||||||||||||||||||||||||||||||||||||:|||||||||||||:|||||||
Db        780 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMS 839

Qy        833 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI 892
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI 899

Qy        893 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 952
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 959

Qy        953 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1012
              |||||||||||||||||||||||||||||||||||||||||||||:||:||||||||||||
Db        960 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDDDM 1019

Qy       1013 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1072
              |||||||||||||||| ||| || |   | ||||||||||||:|||||||||  ||||
Db       1020 GDLVDAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPPRSP 1079

Qy       1073 LAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS 1132
              |||||||||||||||| ||  |||||  || |||||||||||:||| |||||||| ||
Db       1080 LAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPLACS 1139

Qy       1133 PQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENP 1192
              |||||||| :|:|||| |||||  |||||||| ||||||||||||||||||||||||||
Db       1140 PQPEYVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAVENP 1199

Qy       1193 EYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1252
              ||| |: | |  ||| |||||||||||||||:  |:| ||| |:|||||||||||||||
Db       1200 EYLVPREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGLDVP 1259

Qy       1253 V 1253
              |
Db       1260 V 1260
```

RESULT 3
I48161
p-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishi
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MUID:94193007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

Query Match 85.1%; Score 5801.5; DB 2; Length 1254;
Best Local Similarity 84.8%; Pred. No. 1.2e-231;
Matches 1068; Conservative 66; Mismatches 114; Indels 11; Gaps 3;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| | |||||||| |||: ||||||||||||||||||||::|||||||||||||
Db          1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:|||:||| |||||||||||||| ||||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||:| |  || :| |||||||||||||||||||: |||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                  |||||||| ||:| || : ||| | ||||:|| ||:         :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300
```

```
Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||:|||||||||||:|||||||||||||||||| ||||||||||:::: 
Db        301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||:|:|  ||| ||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||:|||||:||:||||| |||||| |||||||||||||| ||| ||||||||||
Db        421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||: |  |:||  :  ||: ||| ||||||||||||||| |||||||
Db        481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |:|||| :|||||||| :|||||||||||||| : || | ||||| || |||| |||||||
Db        541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              |||||||||||||||:|||||  |||||||||||||||::|||||||||| |||:: |||
Db        601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              ||| :|:||| ||||         :|:   |:  ::    :||||||||||||||||||
Db        661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              :|||||||||||||||||||||||||||||||:|||||||||||||:||||||||||||
Db        777 LGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              |||||||||||||| |||||||   | ||||||||||:||||:||| || ||||||
Db       1017 LVDAEEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||:|:||||| || ||:   | ||||||||||:|||:||||||||| ||||
Db       1077 PSEGAGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLACSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||:||||||   |||||  |||||||||| ||||||||||||||| ||||||||||
Db       1137 PEYVNQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              | |:||:| ||| |||  |||||||||||| |||:||:||:|||||||||||||||||
Db       1197 LVPRGGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDVPV 1254
```

RESULT 4
GQHUE
epidermal growth factor receptor precursor - human
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Homo sapiens (man)
C;Date: 15-Nov-1984 #sequence_revision 27-Nov-1985 #text_change 11-Jun-1999
C;Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33

R;Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, rg, P.H.
Nature 309, 418-425, 1984
A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expressi
A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210 <ULL>
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epiderma
A;Reference number: A25772; MUID:85270438; PMID:2991899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identificat
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript ter
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 'RCAWRRA',150-187,'KSVIQAV',195,'M',197,'A',199-222,'S',224-304,'RA',30
',798-799,'TD',802-811,'R',813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF
R;Lin, C.R.; Chen, W.S.; Kruiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verm
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30,'S',32-51;454-467 <WEB>
R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A;Title: Identification of residues in the nucleotide binding site of the epidermal
A;Reference number: A60143; MUID:85182650; PMID:2985580
A;Accession: A60143
A;Molecule type: protein
A;Residues: 740-744,'X',746-747 <RUS>
R;Mroczkowski, B.; Mosig, G.; Cohen, S.

```
Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and superco
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity
A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain r
A;Reference number: A33331; MUID:90003233; PMID:2790960
A;Contents: annotation; internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:120610; OMIM:131550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1210/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <EE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <EE2>
F;646-668/Domain: transmembrane #status predicted <TMM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-726/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F;745/Active site: Lys #status experimental

  Query Match             45.0%;  Score 3069;  DB 1;  Length 1210;
  Best Local Similarity   49.1%;  Pred. No. 3e-119;
  Matches  621;  Conservative  175;  Mismatches  361;  Indels  108;  Gaps   24;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
Db        127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy        189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
Db        184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy        248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
Db        244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy        308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
Db        304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy        367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
Db        362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy        427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
Db        422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy        487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
Db        482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCKL 541

Qy        547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606
Db        542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy        607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG---ILL 663
Db        602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLLL 658

Qy        664 VVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLTPSGAMPNQAQMRILKETELRKVK 722
Db        659 VVALGI--GLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIK 716

Qy        723 VLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSR 782
Db        717 VLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCR 776

Qy        783 LLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHR 842
Db        777 LLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHR 836

Qy        843 DLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSD 902
Db        837 DLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSD 896

Qy        903 VWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSE 962
Db        897 VWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDAD 956

Qy        963 CRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEY 1021
Db        957 SRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEY 1016

Qy       1022 LVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGS 1081
Db       1017 LIPQQGFF---------------------------------SSPSTSRTPLLSSLSATS 1042

Qy       1082 DVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVN 1139
Db       1043 N--NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------PVPEYIN 1094

Qy       1140 QPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-TPQ 1198
Db       1095 Q-SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNPEYLNTVQ 1143

Qy       1199 GGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTPTAE 1242
Db       1144 ---------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGS-TAE 1193

Qy       1243 NPEYL 1247
Db       1194 NAEYL 1198


RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C;Accession: A53183; A43818; S24942; A28941; S45325; I49643
R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF recep
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:U03425
R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding
A;Reference number: A43818; MUID:91232866; PMID:2030916
A;Accession: A43818
A;Molecule type: mRNA
A;Residues: 1-714 <AVI>
A;Cross-references: GB:X59698
```

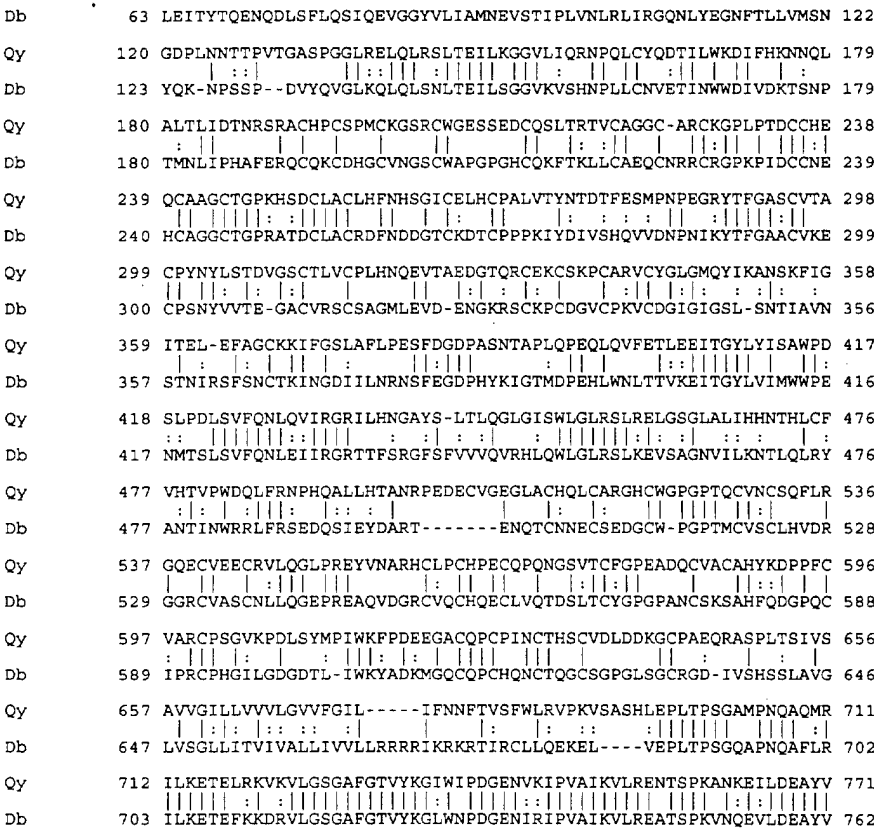
```
R;Eisinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A;Reference number: S24942
A;Accession: S24942
A;Molecule type: mRNA
A;Residues: 969-971,'K',973-1115,'D' <EIS>
A;Cross-references: EMBL:Z12608
R;Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A;Reference number: A28941; MUID:88330814; PMID:3138233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971,'K',973-1210 <VER>
A;Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A;Reference number: I49643; MUID:93126380; PMID:7678348
A;Accession: I49643
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 12-20,22-132 <RES>
A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
C;Genetics:
A;Gene: EGFR
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphopro
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TMM>
F;712-977/Domain: protein kinase homology <KIN>
F;720-728/Region: protein kinase ATP-binding motif
F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

  Query Match             44.6%;  Score 3040;  DB 2;  Length 1210;
  Best Local Similarity   48.9%;  Pred. No. 4.7e-118;
  Matches  622;  Conservative  168;  Mismatches  369;  Indels  114;  Gaps   25;

Qy          11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
               ||  ||    | |     :|| ||  :|    : | |   |: :|  |:|| ||||:||:  |
Db          14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy          69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
                 ||||: |||| |||||| | | ::||: |:|:||  |:|: ||||:| |
Db          74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy         129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
                 | :  |||||  :|:| |||  | |     || ||  ||| |:||     |  |  : |
Db         125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy         185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
                 :   :|  | | |     |||   |:|| ||: :||  |: ||:|  |:|||| |||||
Db         181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy         244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
               ||||: |||| |  |     |:  || |: ||  |::  ||||:|:|||:||  || ||
Db         240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy         304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
               : || |||   |     ||  ||| ::|:||  || :|| |:|:   | ::  |  | ::
Db         300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy         364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
                |  |  | | |  || :| ||  : | || | :|:: :|::||||:| | ||||:  ||
Db         358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy         423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
                 :||::|||   :| :|| : || |: ||||||:|:  |  :|  | :||: :|: |
Db         418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy         483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
                :||  |:|      || | :|      |: ||:   |||| |  ||:|    ||:||||
Db         478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy         543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
               :| :|:| |||:|     |: ||||| ||  ::|| |   | |: |||| | | ||  ||:
Db         538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy         603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
               |:  : : : :||: |     |  |  |||: |     :||         : || : :|| |
Db         598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy         663 LVVVLGVVFGILIF---NNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILKETELR 719
               | :|: |  || :|     :       ||        :|||||| |||| :|||||| :
Db         657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFK 715

Qy         720 KVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPY 779
               |:|||||||||||||:|||:|| |||||||| ||| |||||||||||||||| | :|:
Db         716 KIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPH 775

Qy         780 VSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRL 839
               | |||||||||||||:|||||||||||:|||:: :||| |||||:||||||:|||| ||
Db         776 VCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRL 835

Qy         840 VHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTH 899
               |||||||||||||:| ||||||||||:||  :| ||||:|||||||||||||| | :||
Db         836 VHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTH 895

Qy         900 QSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMI 959
               |||||||||||||||||:|||||||| :|  :|||||||||||||||||||||||||||
Db         896 QSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVKCWMI 955

Qy         960 DSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1018
               |:: ||:|||||: |||:|||||||::||| :|  :    || || |||:|:::|| |:|||
Db         956 DADSRPKFRELILEFSQMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVVDA 1015

Qy        1019 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1078
               :|||:|||||                    | ||                    |:|  |
Db        1016 DEYLIPQQGFF----------------NSPST------------------SRTPLLSSLS 1041

Qy        1079 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPE 1136
               | |:   :  :           |    : : |||| ||| :  :  | ::      | ||
Db        1042 ATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------PVPE 1091

Qy        1137 YVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL- 1195
               ||||  |  :|   : |:  :|          :|::   ::      || |||||
Db        1092 YVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNPEYLN 1140

Qy        1196 TPQGGAAPQPHPPPAPSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTP 1239
               | |          |   |  |::   | |          :|        |:   |   ||| |
Db        1141 TAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIFKG-P 1190

Qy        1240 TAENPEYLGLDVP 1252
               |||| ||| :  |
Db        1191 TAENAEYLRVAPP 1203


RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Gallus gallus (chicken)
C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C;Accession: A27720; A00643
```

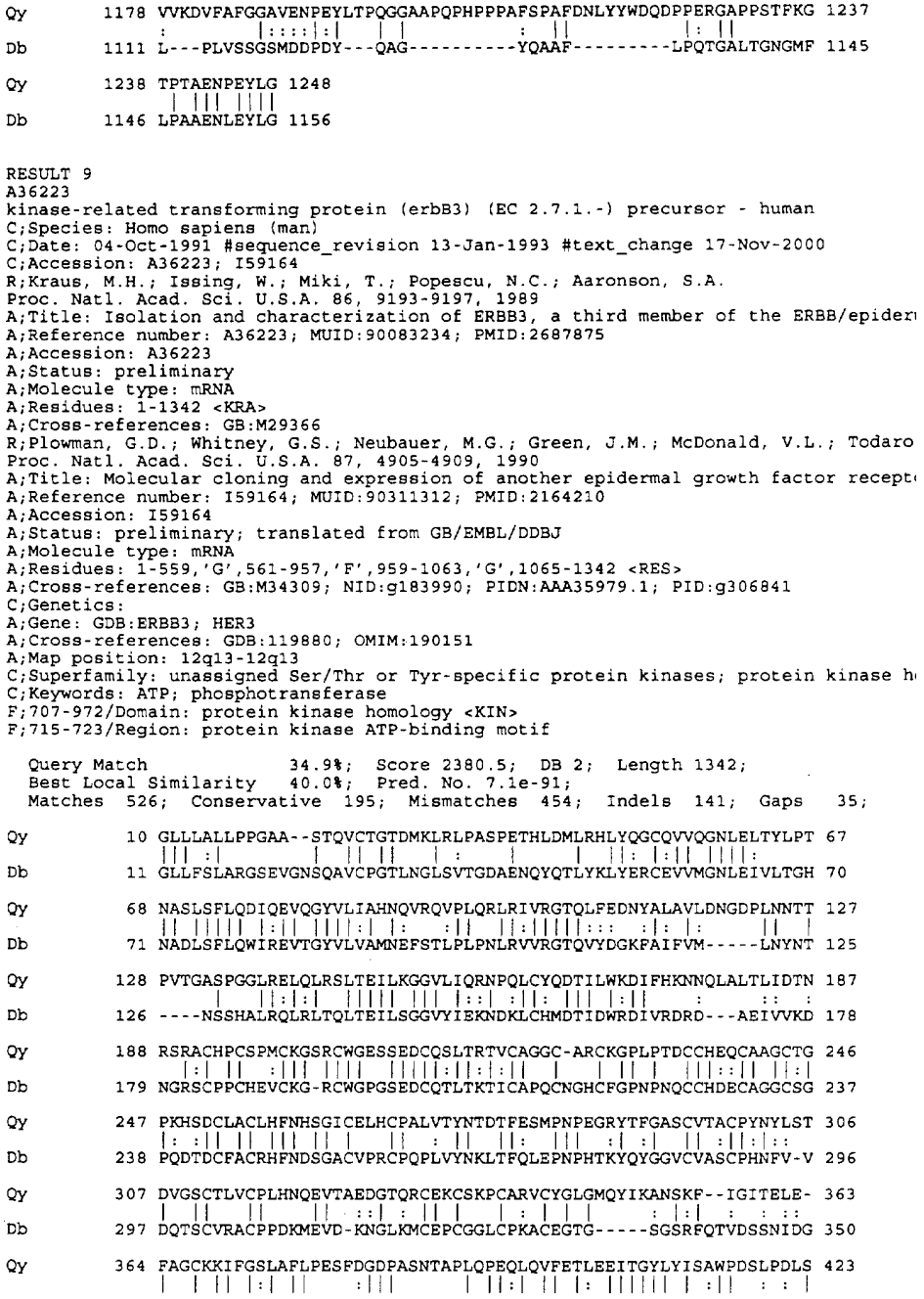
```
R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A;Reference number: A27720; MUID:88261272; PMID:3260329
A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
pecific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TMM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

  Query Match             44.3%;  Score 3017.5;  DB 1;  Length 1223;
  Best Local Similarity   47.9%;  Pred. No. 4e-117;
  Matches  621;  Conservative  171;  Mismatches  358;  Indels  147;  Gaps   27;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |   :| ||  | |         :|| ||: ||       | |    |: :|  |:||  |||
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  |:||: |||| |||||  | |  :||: |:|:||  |:::::|||||
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
               :| |          ||||| :: |:||| ||| |  ||:||  ||:|| ||   : :  |
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              |::|  :   :|  | | |     |||   ::||:||: :||  |: ||:| :|:|||| |
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: ||||||  |      |:  || || ||  |::  ||||:|:|||:|| |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||   |       ||  |:| ::|:||   |::|| |:|:  :|      |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  | || | :: || :| ||  :  | || |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
              ||  |:||::|||   :| ||| :  | |  ||||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
```

```
              |: |  ||    |      || :::|  :   |  ||:  ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
              |||:: | :||| |||:      ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIV 655
              ||  ||:||  :   : :||: |    || |  ||| |     :|||     |   ||
Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 SAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLTPSGAMPNQAQMRIL 713
              : || |:| :||:|:  |: :     |    ||         :|||||| |||| :|||
Db        657 AGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRIL 716

Qy        714 KETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 773
              |||| :|||||||||||||||:|||:|| |||||||| ||| ||||||||||||||||
Db        717 KETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMA 776

Qy        774 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 833
              | :|:| |||||||||||||:|||||||||||::||::  :||| |||||:||||||:|
Db        777 SVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWCVQIAKGMNY 836

Qy        834 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 893
              ||: |||||||||||||:| |||||||||||:||  || |||||||||||||||||||
Db        837 LEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPIKWMALESIL 896

Qy        894 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 953
              | :|||||||||||||||||:|||||||| ||  :|||||||||||||||||||||||
Db        897 HRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPICTIDVYMIM 956

Qy        954 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDM 1012
              |||||||:: ||:||||::|||:||||| |::||| :| :   || || |||:|:|::||
Db        957 VKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFYRTLMEEEDM 1016

Qy       1013 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1072
              |:|||:|||| ||||                | ||                  |:|
Db       1017 EDIVDADEYLVPHQGFF----------------NSPST------------------SRTP 1042

Qy       1073 L-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGY 1125
              |     | |  : ::  |        :  |  |  : | :|||| |||     |:  ||:
Db       1043 LLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNFLEESIDDGF 1094

Qy       1126 VAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVF-- 1183
              :      | ||||||  : |: ||                  :  :| :  ::
Db       1095 L------PAPEYVNQ--LMPKKPS--------------------TAMVQNQIYNNISLT 1125

Qy       1184 -------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQ------- 1223
                          :  ||:|||||        :  |      |::  || |
Db       1126 AISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFESSPYWIQSGNHQIN 1177

Qy       1224 -DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1252
               | |:          |:     | |||||| :  |
Db       1178 LDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 1214
```

```
RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
```

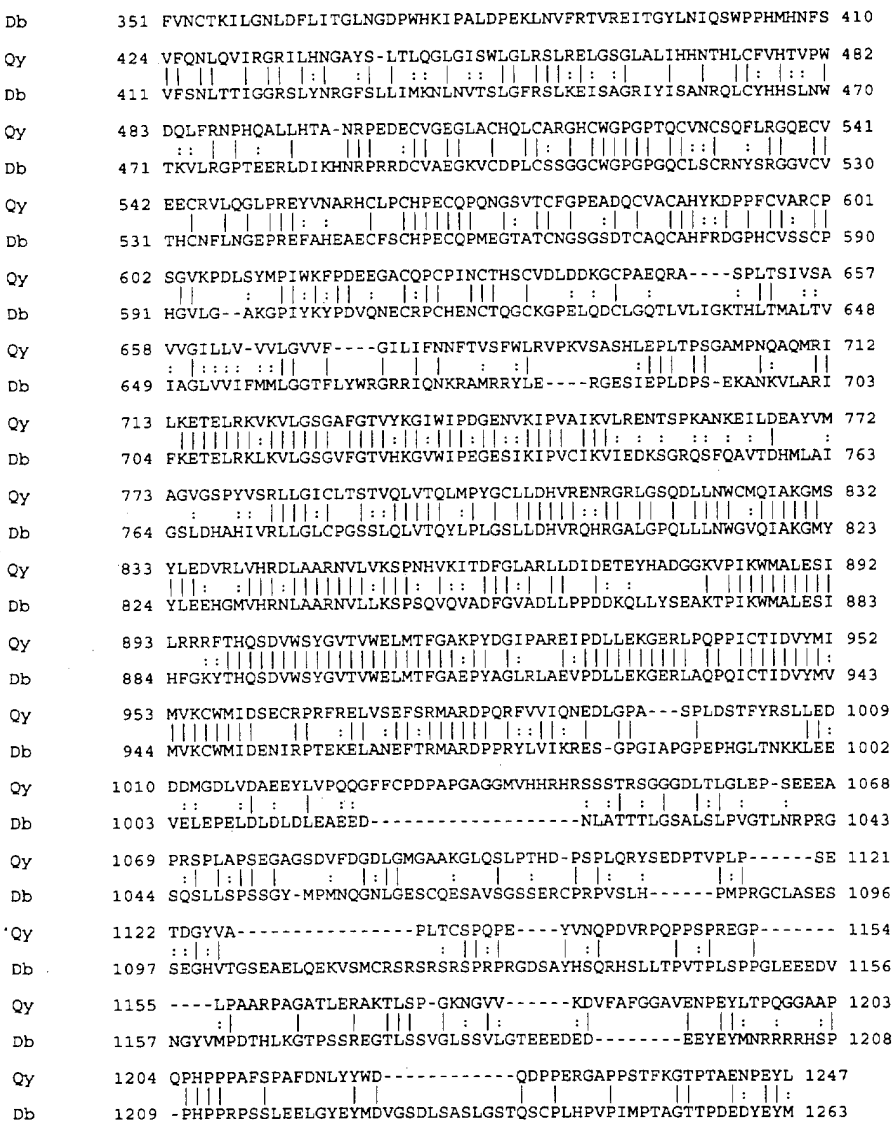
```
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

  Query Match             42.7%;  Score 2908.5;  DB 2;  Length 1308;
  Best Local Similarity   44.4%;  Pred. No. 1.3e-112;
  Matches  600;  Conservative  182;  Mismatches  384;  Indels  185;  Gaps   28;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  |  |||| : ::|| |||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::|||| ||| : :|  ||| ||| |:||         |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |  | | |||| :  ||:|||||| |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||    ||  ||:   |  :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||     ||  |:| : |: |:  | : | |:|   : :          :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :    : ||:| || |: |||:| | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||   | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||    :| ::   ||  : |  ||: |: ||:   |||||| ||::| :| ||: |:|
Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              | :  |  ||:| |    |: | |:|:  ::| :|| ||  | |  :|:|:|| | || :||
Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PAEQRASPL 651
              |::   |:  |:|: | :  | ||  ||| |          |          |  | :||
Db        596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQHAR-TPL 652

Qy        652 TSIVSAVV-GILLVVVLGVVFGILIFNNFTVSFWLRVPKVSASH-------LEPLTPSGA 703
              | :|: |: ::|::|: | : :           : |  |          :|||||||
Db        653 --IAAGVIGGLFILVIVGLTFAVYVRRK-------SIKKKRALRRFLETELVEPLTPSGT 703

Qy        704 MPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 763
              |||||:|||||||::|||||||||||||||:|:|| |||||||:|  |  ||||
Db        704 APNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKANV 763

Qy        764 EILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNW 823
              | :||| :|| :  |:: ||||:|| |:|||||||:|||||::| |::  :||| ||||
Db        764 EFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLLNW 823

Qy        824 CMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVP 883
              |:|||||| |||: |||||||||||||||||||||||||||||: || ||:|||||:|
Db        824 CVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKMP 883

Qy        884 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPP 943
              |||||||  |:|||||||||||||:|||||| |||||| |||||||||||||||||||||
Db        884 IKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQPP 943

Qy        944 ICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDSTF 1002
              ||||||||:|||||||:: ||:|:|| :|||||||||||::|||  :|  :    || || |
Db        944 ICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDSKF 1003

Qy       1003 YRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLE 1062
              :::|:::|: |::|||||||| | |  | |       :   | |  | ||      :|
Db       1004 FQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIGHS 1051

Qy       1063 PSEEEAPRS--------------------------PLAP-SEGAGSDVFDGDLGMGAA 1093
              |     | |                          | || ::|| :::||      |
Db       1052 PPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNGTL 1111

Qy       1094 KGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQ 1146
              :  :    : |  |||| |||| |         : :||: |:   |: ||:|  :
Db       1112 RKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE---- 1167

Qy       1147 PPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPH 1206
                    | |  : |               ||| ::       |::||||       |
Db       1168 -----ENPFVSRR--------------KNGDLQ-------ALDNPEYHNASNG------ 1194

Qy       1207 PPPA--------------------------------FSPAFDNLYYWDQDPPERGA-- 1230
              || |                               ||||  ||:   | |
Db       1195 PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRSTLQ 1254

Qy       1231 PPSTFKGTPT--------------AENPEYL 1247
                |   :   |              |||||||
Db       1255 HPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transfo
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueler, W.; Raulf, F.; Telling, A.; Ro
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueler, W.; Schartl, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025,'N',1027-1098,'A',1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

  Query Match             38.1%;  Score 2598;  DB 1;  Length 1166;
  Best Local Similarity   44.1%;  Pred. No. 7e-100;
  Matches  561;  Conservative  168;  Mismatches  394;  Indels  148;  Gaps   31;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              |||  :  |||  |       ||     :|| ||  :: :    : |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||   |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: | |: |
```



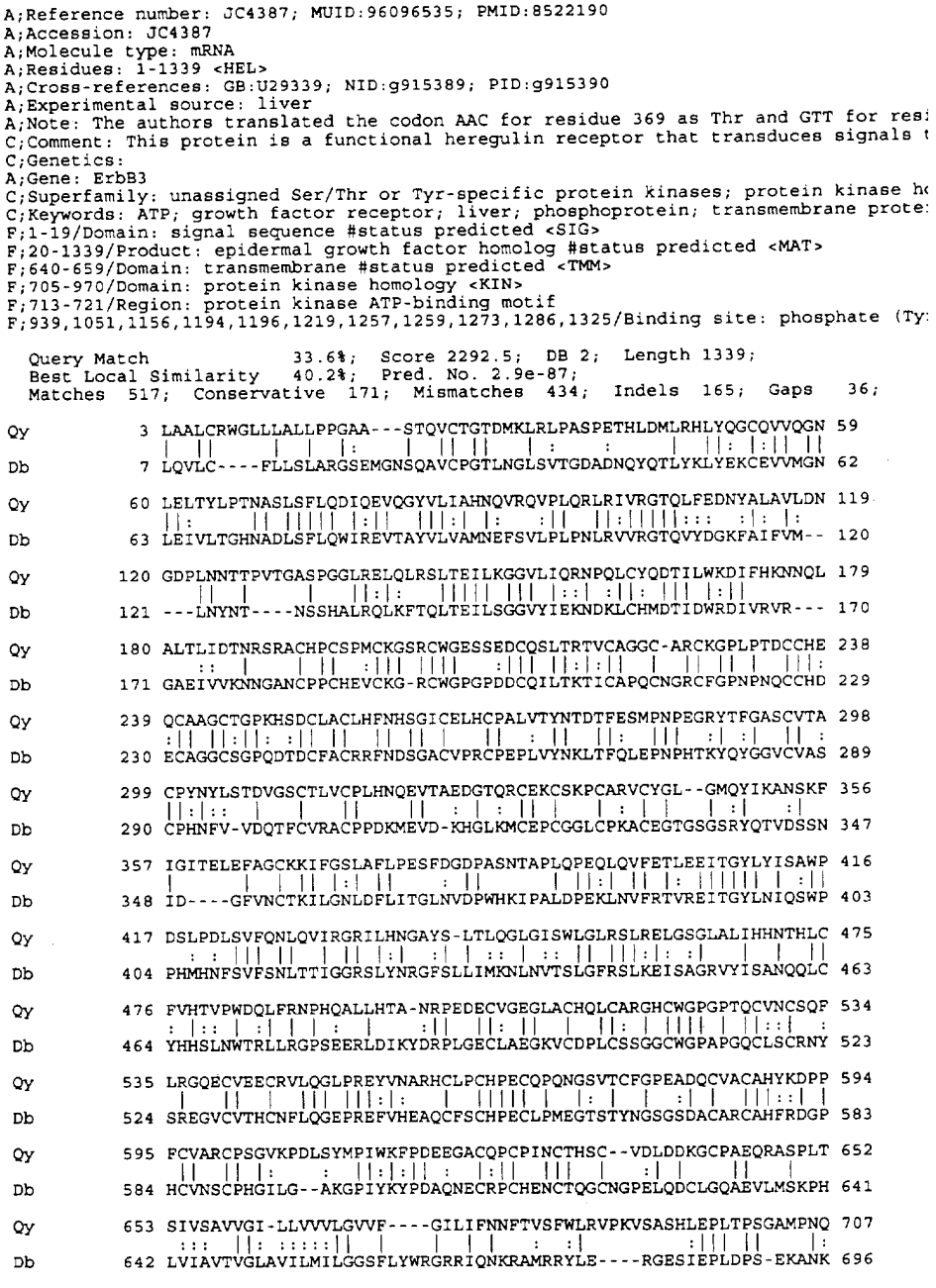


RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C;Accession: A36223; I59164
R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epider
A;Reference number: A36223; MUID:90083234; PMID:2687875
A;Accession: A36223
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1342 <KRA>
A;Cross-references: GB:M29366
R;Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A;Title: Molecular cloning and expression of another epidermal growth factor recept
A;Reference number: I59164; MUID:90311312; PMID:2164210
A;Accession: I59164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-559,'G',561-957,'F',959-1063,'G',1065-1342 <RES>
A;Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C;Genetics:
A;Gene: GDB:ERBB3; HER3
A;Cross-references: GDB:119880; OMIM:190151
A;Map position: 12q13-12q13
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; phosphotransferase
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 34.9%; Score 2380.5; DB 2; Length 1342;
Best Local Similarity 40.0%; Pred. No. 7.1e-91;
Matches 526; Conservative 195; Mismatches 454; Indels 141; Gaps 35;

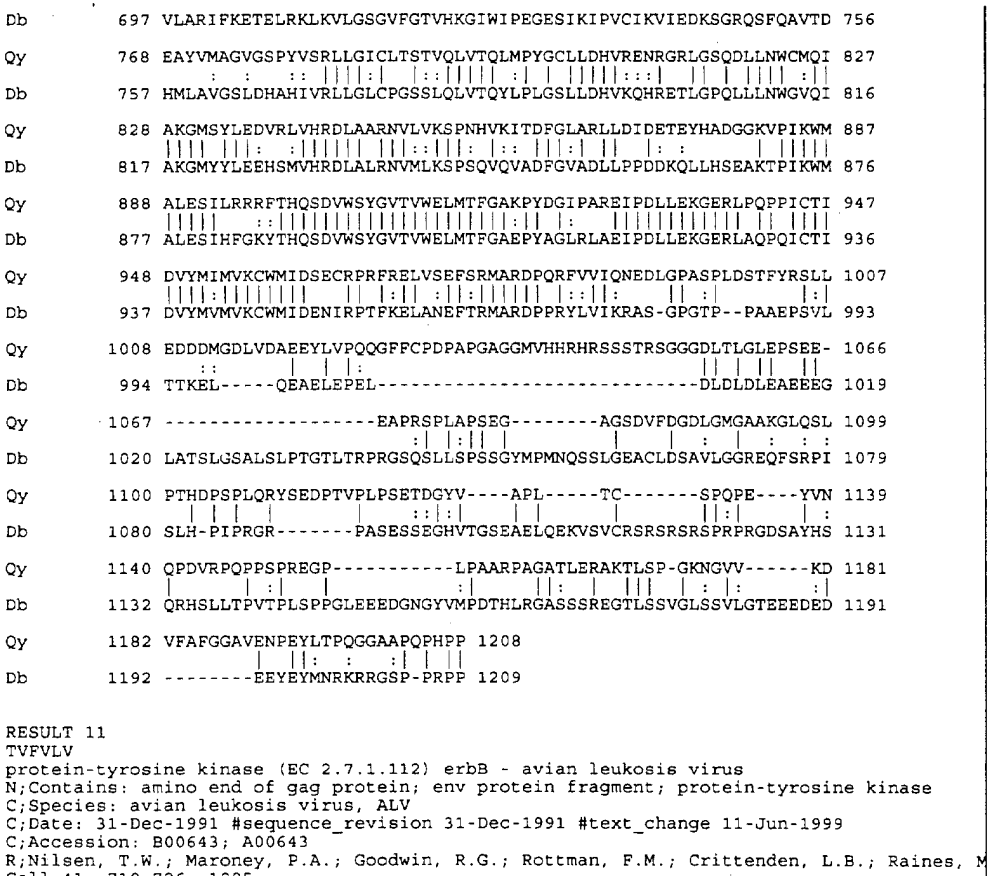


RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.



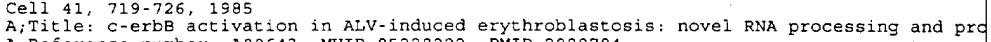
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res
C;Comment: This protein is a functional heregulin receptor that transduces signals
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TMM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Ty

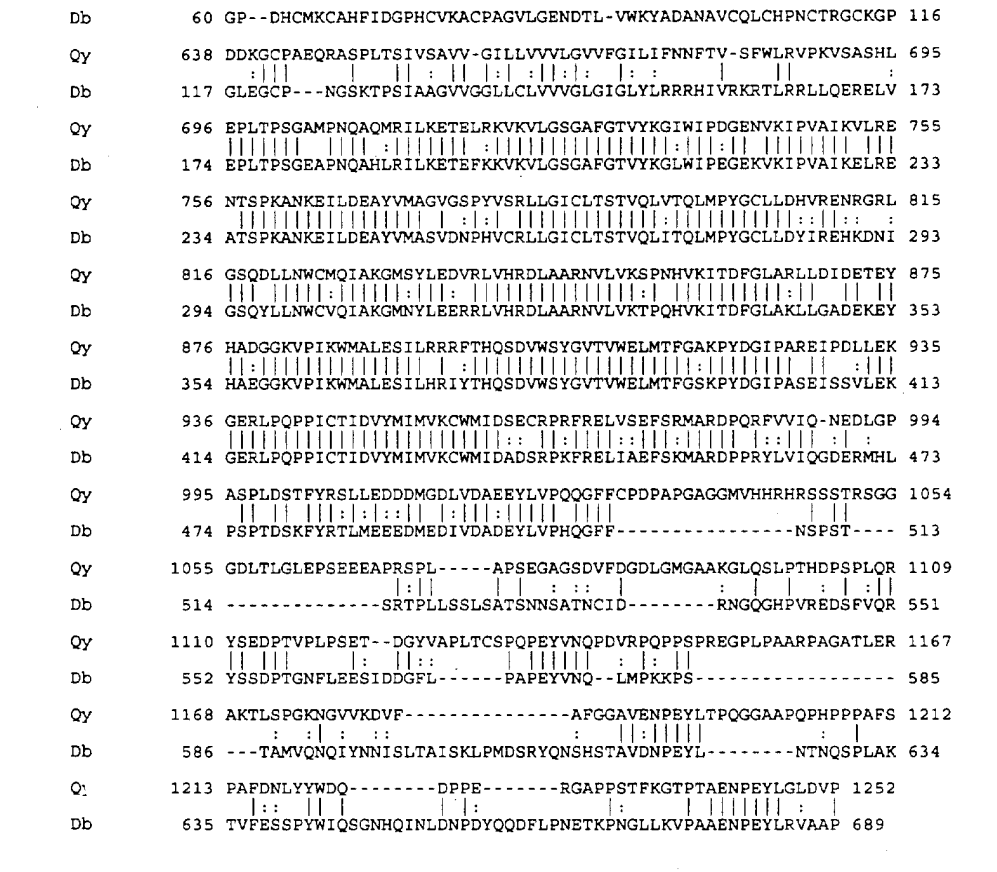
Query Match 33.6%; Score 2292.5; DB 2; Length 1339;
Best Local Similarity 40.2%; Pred. No. 2.9e-87;
Matches 517; Conservative 171; Mismatches 434; Indels 165; Gaps 36;



RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

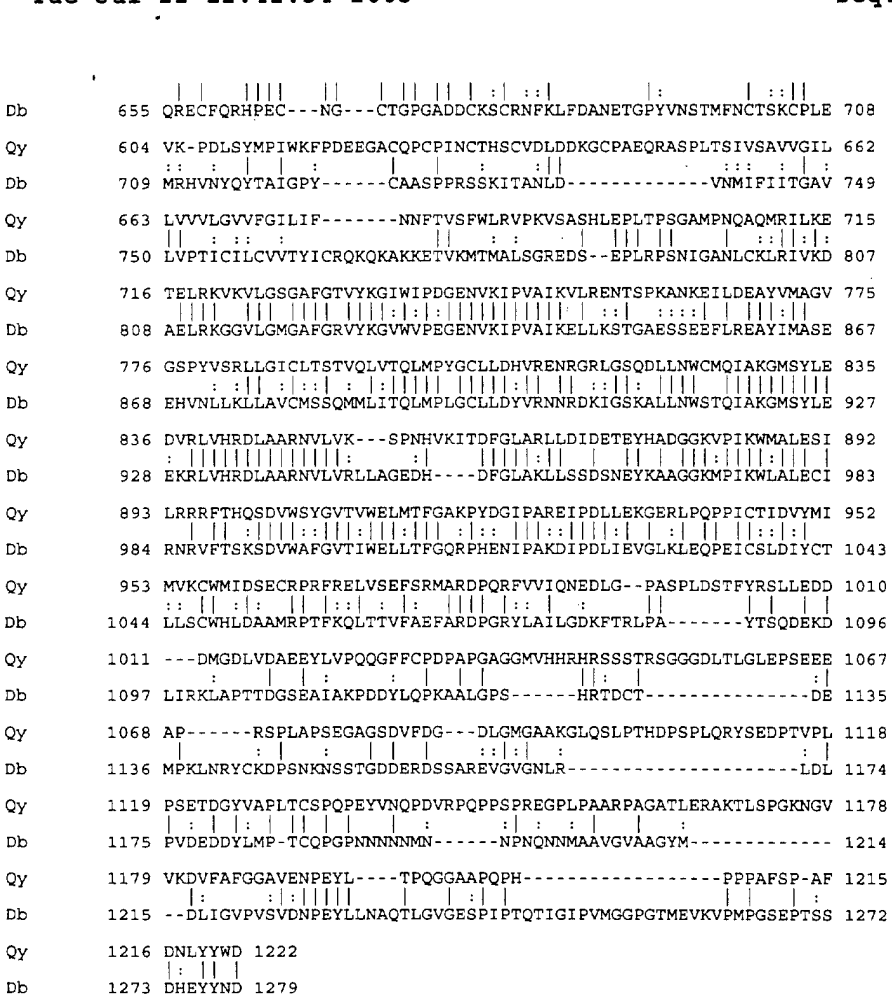
  Query Match 24.8%; Score 1689.5; DB 1; Length 698;
  Best Local Similarity 50.8%; Pred. No. 1e-62;
  Matches 363; Conservative 78; Mismatches 149; Indels 125; Gaps 18;





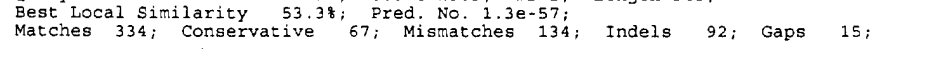
RESULT 12
TVYUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Mart
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specif

```
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

  Query Match             23.9%;  Score 1626;  DB 1;  Length 604;
  Best Local Similarity   50.7%;  Pred. No. 3.5e-60;
  Matches  349;  Conservative   74;  Mismatches  140;  Indels  126;  Gaps   16;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLTPSGAM 704
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEA 118

Qy        705 PNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 764
Db        119 PNQAHLRILKETEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKELREATSPKANKE 178

Qy        765 ILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWC 824
Db        179 ILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWC 238

Qy        825 MQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPI 884
Db        239 VQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPI 298

Qy        885 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 944
Db        299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy        945 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1003
Db        359 CTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy       1004 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1063
Db        419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------ 449

Qy       1064 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1118
Db        450 -----SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNF 496

Qy       1119 PSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKN 1176
Db        497 LEESIDDGFL------PAPEYVNQ--LMPKKPSTAM------------------------ 524

Qy       1177 GVVKDVFAF------------------GGAVENPEYLTPQGGAAPQPHPPPAFSPAFDN 1217
Db        525 -VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFES 575

Qy       1218 LYYWDQDPPERGAPPSTFKGTPTAENPEY 1246
Db        576 SPYWIQSGNHQ----------INLDNPDY 594


RESULT 13
GQFFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <LIV>
A;Cross-references: EMBL:K03054
R;Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth fact
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021
A;Molecule type: DNA
A;Residues: 'A',832-866,'V',868-943,'QTPSLVK' <WAD>
A;Cross-references: EMBL:X02293; NID:g7922; PIDN:CAA26157.1; PID:g929565
C;Comment: This sequence is tentative because the introns have not been identified.
C;Genetics:
A;Gene: FlyBase:Egfr
A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2 57F
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TMM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

  Query Match             23.5%;  Score 1599.5;  DB 1;  Length 1330;
  Best Local Similarity   29.3%;  Pred. No. 9.4e-59;
  Matches  406;  Conservative  177;  Mismatches  415;  Indels  389;  Gaps   40;

Qy         80 VQGYVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASP 134
Db         38 ITNYIVIGLDLIPCTLSYRLQIIRGRTLFSLSVEEEKYALFV----------------TY 81

Qy        135 GGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHP 194
Db         82 SKMYTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPK 141

Qy        195 CSPMCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDC 252
Db        142 CHESCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDC 200

Qy        253 LACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCT 312
Db        201 IACKNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACV 259

Qy        313 LVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAG 366
Db        260 RSCPQDKMDKGGE-----CVPCNGPCPKTC--------------PGVTVLHAGNIDSFRN 300

Qy        367 CKKIFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDS 418
Db        301 CTVIDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQ 360

Qy        419 LPDLSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFV 477
Db        361 FRNLSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYV 420

Qy        478 HTVPWDQLFRNPHQALLHTANRPEDEC---------------------------------- 504
Db        421 SNIRWPAIQKEPEQKVWVNENLRADLCGKFLTILISVQHNIIMHIFAICREKWNHLLGSV 480

Qy        505 ------------------------------------------------------------ 504
Db        481 QRGRLLGSWHGSVPYLQELQFQWHLHRRLWLYIQVSINSTQDKSNEHQLTDACYSPSVPT 540

Qy        505 ------------------------------------------------------------VG 506
Db        541 SLTIERARYAIQSAGLAMELEQITARSASMRHSKTLPAEGRQVPRWVFLGVCASARAGIA 600

Qy        507 EGLA-------CHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYV---N 556
Db        601 EPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCETEC------PADHYTDEE 654

Qy        557 ARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYK------DPPF-------CVARCPSG 603
```



RESULT 14
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
R;Vennstroem, B.
submitted to the EMBL Data Library, March 1993
A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X12707
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

RESULT 15
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus m
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif



```
Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHL 695
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELV 114

Qy        696 EPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRE 755
Db        115 EPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELRE 174

Qy        756 NTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRL 815
Db        175 ATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNI 234

Qy        816 GSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEY 875
Db        235 GSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQLGADEKEY 294

Qy        876 HADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 935
Db        295 HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354

Qy        936 GERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGP 994
Db        355 GERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHL 414

Qy        995 ASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGG 1054
Db        415 PSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST---- 454

Qy       1055 GDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQR 1109
Db        455 --------------SRTPLLSSLSATSNNSATNCIDRNGG------------H------- 481

Qy       1110 YSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERA 1168
Db        482 ---------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAISKL 524

Qy       1169 KTLSPGKNGVVKDVFAFGGAVENPEYL 1195
Db        525 PMDSRYQN-------SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:27:42
Job time : 30.25 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.625 Seconds
(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-653-675-14
Perfect score: 6814
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6604	96.9	1255	1	ERB2_HUMAN	P04626 homo sapien
2	5817	85.4	1257	1	ERB2_RAT	P06494 rattus norv
3	5801.5	85.1	1254	1	ERB2_MESAU	Q60553 mesocricetu
4	3067	45.0	1210	1	EGFR_HUMAN	P00533 homo sapien
5	3041	44.6	1210	1	EGFR_MOUSE	Q01279 mus musculu
6	2908.5	42.7	1308	1	ERB4_HUMAN	Q15303 homo sapien
7	2889	42.4	1308	1	ERB4_RAT	Q62956 rattus norv
8	2614.5	38.4	1167	1	XMRK_XIPMA	P13388 xiphophorus
9	2389.5	35.1	1342	1	ERB3_HUMAN	P21860 homo sapien
10	2315.5	34.0	1339	1	ERB3_RAT	Q62799 rattus norv
11	1924	28.2	1426	1	EGFR_DROME	P04412 drosophila
12	1672.5	24.5	634	1	ERBB_ALV	P00534 avian leuko
13	1626	23.9	604	1	ERBB_AVIER	P00535 avian eryth
14	1553	22.8	540	1	ERBB_AVIEU	P11273 avian eryth
15	1518	22.3	703	1	EGFR_CHICK	P13387 gallus gall
16	1287	18.9	1323	1	LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1	ERB2_MOUSE	P70424 mus musculu
18	724	10.6	1363	1	ILPR_BRALA	O02466 branchiosto
19	706	10.4	1383	1	INSR_RAT	P15127 rattus norv
20	705.5	10.4	1372	1	INSR_MOUSE	P15208 mus musculu
21	705	10.3	1300	1	IRR_MOUSE	Q9wtl4 mus musculu
22	703	10.3	1382	1	INSR_HUMAN	P06213 homo sapien
23	692.5	10.2	1300	1	IRR_CAVPO	P14617 cavia porce
24	685	10.1	1297	1	IRR_HUMAN	P14616 homo sapien
25	685	10.1	1477	1	HTK7_HYDAT	Q25197 hydra atten
26	680	10.0	1607	1	MIPR_LYMST	Q25410 lymnaea sta
27	638	9.4	1367	1	IG1R_HUMAN	P08069 homo sapien
28	625	9.2	1373	1	IG1R_MOUSE	Q60751 mus musculu
29	625	9.2	2146	1	INSR_DROME	P09208 drosophila
30	622	9.1	1390	1	INSR_AEDAE	Q93105 aedes aegyp
31	621.5	9.1	1370	1	IG1R_RAT	P24062 rattus norv
32	604	8.9	1114	1	RET_HUMAN	P07949 homo sapien
33	592	8.7	987	1	EPB4_HUMAN	P54760 homo sapien
34	584.5	8.6	984	1	EPB1_CHICK	Q07494 gallus gall
35	576.5	8.5	984	1	EPB1_RAT	P09759 rattus norv
36	573	8.4	902	1	EPBB_XENLA	Q91736 xenopus lae
37	572.5	8.4	977	1	EPA2_MOUSE	Q03145 mus musculu
38	572	8.4	987	1	EPB4_MOUSE	P54761 mus musculu
39	570.5	8.4	984	1	EPB1_HUMAN	P54762 homo sapien
40	569	8.4	757	1	HT16_HYDAT	P53356 hydra atten
41	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
42	568.5	8.3	1053	1	FAK1_CHICK	Q00944 gallus gall
43	567	8.3	1052	1	FAK1_MOUSE	P34152 mus musculu
44	565.5	8.3	985	1	EPBA_XENLA	Q91571 xenopus lae
45	562	8.2	1055	1	FAK1_RAT	O35346 rattus norv

ALIGNMENTS

```
RESULT 1
ERB2_HUMAN
ID   ERB2_HUMAN     STANDARD;      PRT;  1255 AA.
AC   P04626;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE   surface receptor HER2) (MLN 19).
GN   ERBB2 OR HER2 OR NGL OR NEU.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86118663; PubMed=3003577;
RA   Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA   Saito T., Toyoshima K.;
RT   "Similarity of protein encoded by the human c-erb-B-2 gene to
RT   epidermal growth factor receptor.";
RL   Nature 319:230-234(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86070181; PubMed=2999974;
RA   Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA   McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA   Francke U., Levinson A., Ullrich A.;
RT   "Tyrosine kinase receptor with extensive homology to EGF receptor
RT   shares chromosomal location with neu oncogene.";
RL   Science 230:1132-1139(1985).
RN   [3]
RP   SEQUENCE OF 737-1031 FROM N.A.
RX   MEDLINE=86016729; PubMed=2995967;
RA   Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT   "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT   c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT   human salivary gland adenocarcinoma.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN   [4]
RP   VARIANTS VAL-654 AND VAL-655.
RX   MEDLINE=93194196; PubMed=8095488;
RA   Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT   "Characterization of a new allele of the human ERBB2 gene by allele-
RT   specific competition hybridization.";
RL   Genomics 15:426-429(1993).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
```

```
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
CC       POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC       OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC       ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11767; AAA35808.1; -.
DR   EMBL; M11761; AAA35808.1; JOINED.
DR   EMBL; M11762; AAA35808.1; JOINED.
DR   EMBL; M11763; AAA35808.1; JOINED.
DR   EMBL; M11764; AAA35808.1; JOINED.
DR   EMBL; M11765; AAA35808.1; JOINED.
DR   EMBL; M11766; AAA35808.1; JOINED.
DR   EMBL; M11730; AAA75493.1; -.
DR   EMBL; M12036; AAA35978.1; -.
DR   EMBL; X03363; CAA27060.1; -.
DR   PIR; A25491; A25491.
DR   PIR; A24571; A24571.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3430; ERBB2.
DR   MIM; 164870; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Polymorphism.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   1255       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN      22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   653    675       POTENTIAL.
FT   DOMAIN     676   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     720    987       PROTEIN KINASE.
FT   NP_BIND    726    734       ATP (BY SIMILARITY).
FT   BINDING    753    753       ATP (BY SIMILARITY).
FT   ACT_SITE   845    845       BY SIMILARITY.
FT   DISULFID   195    204       BY SIMILARITY.
FT   DISULFID   199    212       BY SIMILARITY.
FT   DISULFID   220    227       BY SIMILARITY.
FT   DISULFID   224    235       BY SIMILARITY.
FT   DISULFID   236    244       BY SIMILARITY.
FT   DISULFID   240    252       BY SIMILARITY.
FT   DISULFID   255    264       BY SIMILARITY.
FT   DISULFID   268    295       BY SIMILARITY.
FT   DISULFID   299    311       BY SIMILARITY.
FT   DISULFID   315    331       BY SIMILARITY.
FT   DISULFID   334    338       BY SIMILARITY.
FT   DISULFID   511    520       BY SIMILARITY.
FT   DISULFID   515    528       BY SIMILARITY.
FT   DISULFID   531    540       BY SIMILARITY.
FT   DISULFID   544    560       BY SIMILARITY.
FT   DISULFID   563    576       BY SIMILARITY.
FT   DISULFID   567    584       BY SIMILARITY.
FT   DISULFID   587    596       BY SIMILARITY.
FT   DISULFID   600    623       BY SIMILARITY.
FT   DISULFID   626    634       BY SIMILARITY.
FT   DISULFID   630    642       BY SIMILARITY.
FT   MOD_RES   1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    654    654       I -> V.
FT                               /FTId=VAR_004077.
FT   VARIANT    655    655       I -> V.
FT                               /FTId=VAR_004078.
FT   CONFLICT  1170   1170       P -> A (IN REF. 2).
SQ   SEQUENCE  1255 AA;  137909 MW;  39E9DFDA04DCF962 CRC64;

  Query Match             96.9%;  Score 6604;  DB 1;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||         :|:   |: ::    :||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716
```

```
Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 2
ERB2_RAT
ID   ERB2_RAT       STANDARD;      PRT;  1257 AA.
AC   P06494;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE   receptor-related protein).
GN   ERBB2 OR NEU.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=86118662; PubMed=3945311;
RA   Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT   "The neu oncogene encodes an epidermal growth factor receptor-related
RT   protein.";
RL   Nature 319:226-230(1986).
RN   [2]
RP   SEQUENCE OF 852-905 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   STRUCTURE BY NMR OF 650-668.
RX   MEDLINE=92155181; PubMed=1346763;
RA   Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,
RA   Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT   "Three dimensional structure of the transmembrane region of the proto-
RT   oncogenic and oncogenic forms of the neu protein.";
RL   EMBO J. 11:43-48(1992).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC       THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03362; CAA27059.1; ALT_INIT.
DR   PIR; A24562; TVRTNU.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1257       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    677       POTENTIAL.
FT   DOMAIN      678   1257       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      159    369       CYS-RICH.
FT   DOMAIN      473    646       CYS-RICH.
FT   DOMAIN      722    989       PROTEIN KINASE.
FT   NP_BIND     728    736       ATP (BY SIMILARITY).
FT   BINDING     755    755       ATP (BY SIMILARITY).
FT   ACT_SITE    847    847       BY SIMILARITY.
FT   DISULFID    196    205       BY SIMILARITY.
FT   DISULFID    200    213       BY SIMILARITY.
FT   DISULFID    221    228       BY SIMILARITY.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    237    245       BY SIMILARITY.
FT   DISULFID    241    253       BY SIMILARITY.
FT   DISULFID    256    265       BY SIMILARITY.
FT   DISULFID    269    296       BY SIMILARITY.
FT   DISULFID    300    312       BY SIMILARITY.
FT   DISULFID    316    332       BY SIMILARITY.
FT   DISULFID    335    339       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    578       BY SIMILARITY.
```

```
FT   DISULFID    569    586       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    602    625       BY SIMILARITY.
FT   DISULFID    628    636       BY SIMILARITY.
FT   DISULFID    632    644       BY SIMILARITY.
FT   MOD_RES    1141   1141       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1250   1250       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    260    260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     661    661       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1257 AA;  138831 MW;  6129264583011402 CRC64;

  Query Match             85.4%;  Score 5817;  DB 1;  Length 1257;
  Best Local Similarity   85.2%;  Pred. No. 5.1e-306;
  Matches 1075;  Conservative   57;  Mismatches  117;  Indels   12;  Gaps    4;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL  60
              ||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL  60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG  120
              ||||:| |||||||||||||||:|||||||::|||||||||||||||| ||||||||
Db         61 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR  120

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL  179
              || :|    | | :| |||||||||||||||||||: |||||||| :||||:| |||||
Db        121 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL  180

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ  239
              |  |||||||||  ||:| || : ||||| |||| || |:|  |||||||  |||||||||
Db        181 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ  240

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC  299
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||| |
Db        241 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC  300

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI  359
              |||||||:|||||||| :|||||||||||||||||||||||||||::::
Db        301 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD  360

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL  419
                || |||||||||||||||||||:|  |||:|||||||||||||||||||||||||||||
Db        361 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL  420

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT  479
              ||||||||::|||||||:|||||||||||||  |||||||||||||||| | |||||||
Db        421 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT  480

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ  538
              |||||||||||||||: ||||:: ||  || |: ||| |||||||||||||| |||||
Db        481 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ  540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA  598
              |||||||| :|||||||: : |||||||||||| | |||| |||||  |||||||   |||
Db        541 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA  600

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAV  658
              |||||||||||||||||:||||| |||||||||||||||::|||||||||||:| |:: |
Db        601 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLLDERGCPAEQRASPVTFIIATV  660

Qy        659 VGILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRI  712
              ||:|| ::| || ||||          :|:   |:  ::    :|||||||||||||||
Db        661 VGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRI  716

Qy        713 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM  772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM  776

Qy        773 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMS  832
              |||||||||||||||||||||||||||||||:|||||||||||||:|||||||
Db        777 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMS  836

Qy        833 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI  892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI  896

Qy        893 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI  952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI  956

Qy        953 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM  1012
              ||||||||||||||||||||||||||||||||||||||||||:||:|||||||||||||
Db        957 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDDDM  1016

Qy       1013 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP  1072
              |||||||||||||||||| ||| || |  | ||||||||||:|||||||||  ||||
Db       1017 GDLVDAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPPRSP  1076

Qy       1073 LAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS  1132
              |||||||||||||||| ||  ||||||  || ||||||||||:||| |||||||| ||
Db       1077 LAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPLACS  1136

Qy       1133 PQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENP  1192
              |||||||| :|:||||   |||||  |||||||| ||||||||||||||||||||||||
Db       1137 PQPEYVNQSEVQPQPPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAVENP  1196

Qy       1193 EYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP  1252
              ||| |: | |  ||| |||||||||||||||:  |:| ||| |:|||||||||||||||
Db       1197 EYLVPREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGLDVP  1256

Qy       1253 V 1253
              |
Db       1257 V 1257


RESULT 3
ERB2_MESAU
ID   ERB2_MESAU     STANDARD;      PRT;  1254 AA.
AC   Q60553;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN   ERBB2 OR NEU.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Nerve;
RX   MEDLINE=94193007; PubMed=7908275;
RA   Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA   Yamazaki Y., Ishikawa T.;
RT   "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL   Gene 140:251-255(1994).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D16295; BAA03801.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1254       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1254       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      158    368       CYS-RICH.
FT   DOMAIN      472    644       CYS-RICH.
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1247   1247       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     658    658       V -> E (IN ONCOGENIC NEU).
FT   VARIANT     659    659       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1254 AA;  138252 MW;  974C3791C21F2BE1 CRC64;

  Query Match             85.1%;  Score 5801.5;  DB 1;  Length 1254;
  Best Local Similarity   84.8%;  Pred. No. 3.5e-305;
  Matches 1068;  Conservative   66;  Mismatches  114;  Indels   11;  Gaps    3;
```

```
Qy      1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
          ||||| | |||||||| |||: |||||||||||||||||||||::||||||||||||||
Db      1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy     61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
          |||||| ||:|||||||||||||:||||:||| |||||||||||||||| ||||||||
Db     61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy    121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
          |||:| |  || :| |||||||||||||||||||: |||||||||:||||:| ||||||
Db    121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy    181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
             |||||||| ||:| || : ||| | ||||:|| |:        :  ||||||||||
Db    181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy    241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300

Qy    301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
          ||||||:|||||||||:|||||||||||||||| |||||||||:::: 
Db    301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

Qy    361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
             |||||||||||||||||||:|:|  ||| |||||||||||||||||||||||||
Db    361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy    421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
          ||||||||:|||||:||:|||||| |||||| |||||||||||| ||| ||||||||||
Db    421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy    481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
          |||||||||||||||: |  |:||  :  ||: ||| |||||||||||||| |||||||
Db    481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy    541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
          |:|||| :|||||||| :|||||||||||| : || | ||||| || |||| |||||||
Db    541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy    601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
          |||||||||||||||:||||| |||||||||||||||::|||||||||| |||:: |||
Db    601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

Qy    661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
          ||| :|:||| |||||         :|:   |:  ::    :|||||||||||||||||
Db    661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy    715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy    775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
          :|||||||||||||||||||||||||||||||||||:|||||||||||||:|||||||||
Db    777 LGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYL 836

Qy    835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
          |||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db    837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALESILR 896

Qy    895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy    955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
          |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db    957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDMGD 1016

Qy   1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
          |||||||||||||| |||||||  | ||||||||||:||||:||| || ||||||
Db   1017 LVDAEEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRSPLA 1076
```

```
Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
             ||||||||||:|:|||| || ||:   | ||||||||||:|||:|||||||| ||||
Db       1077 PSEGAGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLACSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
             |||||||:||||||   |||||  ||||||||| ||||||||||||||| ||||||||||
Db       1137 PEYVNQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
             | |:||:| ||| |||  |||||||||||||| |||:||:||:|||||||||||||||||
Db       1197 LVPRGGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDVPV 1254
```

RESULT 4
EGFR_HUMAN
ID   EGFR_HUMAN   STANDARD;   PRT; 1210 AA.
AC   P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC   O00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE   protein-tyrosine kinase ErbB-1).
GN   EGFR OR ERBB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=84219729; PubMed=6328312;
RA   Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA   Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA   Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT   "Human epidermal growth factor receptor cDNA sequence and aberrant
RT   expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL   Nature 309:418-425(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=95382957; PubMed=7654368;
RA   Ilekis J.V., Stark B.C., Scoccia B.;
RT   "Possible role of variant RNA transcripts in the regulation of
RT   epidermal growth factor receptor expression in human placenta.";
RL   Mol. Reprod. Dev. 41:149-156(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97078686; PubMed=8918811;
RA   Reiter J.L., Maihle N.J.;
RT   "A 1.8 kb alternative transcript from the human epidermal growth
RT   factor receptor gene encodes a truncated form of the receptor.";
RL   Nucleic Acids Res. 24:4050-4056(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97256547; PubMed=9103388;
RA   Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT   "Expression of a truncated epidermal growth factor receptor-like
RT   protein (TEGFR) in ovarian cancer.";
RL   Gynecol. Oncol. 65:36-41(1997).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC   TISSUE=Placenta;
RX   MEDLINE=21100872; PubMed=11161793;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative EGFR transcripts encoding truncated receptor
RT   isoforms.";
RL   Genomics 71:1-20(2001).
RN   [6]
RP   SEQUENCE OF 575-687 FROM N.A.
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Human and mouse alternative EGFR transcripts encoding only the
RT   extracellular domain of the receptor.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 713-924 FROM N.A.
RX   MEDLINE=84196372; PubMed=6326261;
RA   Lin C.R., Chen W.S., Kruiger W., Stolarsky L.S., Weber W.,
RA   Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT   "Expression cloning of human EGF receptor complementary DNA: gene
RT   amplification and three related messenger RNA products in A431
RT   cells.";
RL   Science 224:843-848(1984).
RN   [8]
RP   SEQUENCE OF 150-962 FROM N.A.
RX   MEDLINE=84245835; PubMed=6330563;
RA   Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA   Roe B.A., Merlino G.T., Pastan I.;
RT   "Human epidermal growth factor receptor cDNA is homologous to a
RT   variety of RNAs overproduced in A431 carcinoma cells.";
RL   Nature 309:806-810(1984).
RN   [9]
RP   SEQUENCE OF 1028-1210 FROM N.A.
RX   MEDLINE=85046483; PubMed=6093780;
RA   Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA   O'Malley B.W.;
RT   "Isolation of an evolutionarily conserved epidermal growth factor
RT   receptor cDNA from human A431 carcinoma cells.";
RL   Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN   [10]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=88217333; PubMed=3329716;
RA   Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA   Waterfield M.D.;
RT   "The human EGF receptor gene: structure of the 110 kb locus and
RT   identification of sequences regulating its transcription.";
RL   Oncogene Res. 1:375-396(1987).
RN   [11]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=91107677; PubMed=1988448;
RA   Haley J.D., Waterfield M.D.;
RT   "Contributory effects of de novo transcription and premature
RT   transcript termination in the regulation of human epidermal growth
RT   factor receptor proto-oncogene RNA synthesis.";
RL   J. Biol. Chem. 266:1746-1753(1991).
RN   [12]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=85270438; PubMed=2991899;
RA   Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT   "Characterization and sequence of the promoter region of the human
RT   epidermal growth factor receptor gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN   [13]
RP   SEQUENCE OF 540.
RA   Kohda D.;
RL   Submitted (SEP-1997) to the SWISS-PROT data bank.
RN   [14]
RP   RECEPTOR ACTIVITY.
RX   MEDLINE=84191554; PubMed=6325948;
RA   Mroczkowski B., Mosig G., Cohen S.;
RT   "ATP-stimulated interaction between epidermal growth factor receptor
RT   and supercoiled DNA.";
RL   Nature 309:270-273(1984).
RN   [15]
RP   PHOSPHORYLATION.
RX   MEDLINE=89278137; PubMed=2543678;
RA   Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA   Howk R., Givol D., Ullrich A., Schlessinger J.;
RT   "All autophosphorylation sites of epidermal growth factor (EGF)

```
RT   receptor and HER2/neu are located in their carboxyl-terminal tails.
RT   Identification of a novel site in EGF receptor.";
RL   J. Biol. Chem. 264:10667-10671(1989).
RN   [16]
RP   CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
RP   ASN-528.
RX   MEDLINE=96398132; PubMed=8962717;
RA   Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
RT   "Analysis of the glycosylation patterns of the extracellular domain of
RT   the epidermal growth factor receptor expressed in Chinese hamster
RT   ovary fibroblasts.";
RL   Growth Factors 13:121-132(1996).
RN   [17]
RP   CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
RP   ASN-603.
RX   MEDLINE=20198209; PubMed=10731668;
RA   Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
RT   "Characterization of the N-oligosaccharides attached to the atypical
RT   Asn-X-Cys sequence of recombinant human epidermal growth factor
RT   receptor.";
RL   J. Biochem. 127:65-72(2000).
RN   [18]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=98225196; PubMed=9556602;
RA   Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
RT   "Disulfide bond structure of human epidermal growth factor receptor.";
RL   J. Biol. Chem. 273:11150-11157(1998).
RN   [19]
RP   REVIEW.
RX   MEDLINE=87297456; PubMed=3039909;
RA   Carpenter G.;
RT   "Receptors for epidermal growth factor and other polypeptide
RT   mitogens.";
RL   Annu. Rev. Biochem. 56:881-914(1987).
CC   -!- FUNCTION: Receptor for EGF, but also for other members of the EGF
CC       family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
CC       EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
CC       involved in the control of cell growth and differentiation.
CC   -!- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
CC       secreted.
CC   -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
CC       truncated isoform/TEGFR, 3/p110 and 4; are produced by
CC       alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
CC       expressed in ovarian cancers.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00588; CAA25240.1; -.
DR   EMBL; U95089; AAB53063.1; -.
DR   EMBL; U48722; AAC50802.1; -.
DR   EMBL; U48723; AAC50804.1; -.
DR   EMBL; U48724; AAC50796.1; -.
DR   EMBL; U48725; AAC50797.1; -.
DR   EMBL; U48726; AAC50798.1; -.

  Query Match             45.0%;  Score 3067;  DB 1;  Length 1210;
  Best Local Similarity   49.1%;  Pred. No. 6.3e-158;
  Matches  621;  Conservative  174;  Mismatches  362;  Indels  108;  Gaps   24;
```

```
Qy       11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db       14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy       69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db       74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy      129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
Db      127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy      189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
Db      184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy      248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
Db      244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy      308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
Db      304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy      367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
Db      362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy      427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
Db      422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy      487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
Db      482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNL 541

Qy      547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606
Db      542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy      607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG---ILL 663
Db      602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLL 658

Qy      664 VVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLTPSGAMPNQAQMRILKETELRKVK 722
Db      659 VVALGI--GLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIK 716

Qy      723 VLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSR 782
Db      717 VLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCR 776

Qy      783 LLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHR 842
Db      777 LLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHR 836

Qy      843 DLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSD 902
Db      837 DLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSD 896

Qy      903 VWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSE 962
Db      897 VWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDAD 956

Qy      963 CRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEY 1021
Db      957 SRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEY 1016

Qy     1022 LVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGS 1081
Db     1017 LIPQQGPF----------------------------------SSPSTSRTPLLSSLSATS 1042
```

```
Qy       1082 DVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVN 1139
              :  :  :      |||| |  : | ||||| |||  |  ::  | ::      | |||:|
Db       1043 N--NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------PVPEYIN 1094

Qy       1140 QPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-TPQ 1198
              |  |  :|    : |:   :|          :| ::   :|       || ||||| | |
Db       1095 Q-SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNPEYLNTVQ 1143

Qy       1199 GGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTPTAE 1242
                        |   :  ||:  :| |         :|       |:   |   |||: |||
Db       1144 ---------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGS-TAE 1193

Qy       1243 NPEYL 1247
              | |||
Db       1194 NAEYL 1198
```

```
RESULT 5
EGFR_MOUSE
ID   EGFR_MOUSE     STANDARD;      PRT;  1210 AA.
AC   Q01279;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=93026370; PubMed=1408137;
RA   Avivi A., Skorecki K., Yayon A., Givol D.;
RT   "Promoter region of the murine fibroblast growth factor receptor 2
RT   (bek/KGFR) gene.";
RL   Oncogene 7:1957-1962(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX   MEDLINE=93126380; PubMed=7678348;
RA   Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT   "Expression of the epidermal growth factor receptor gene is regulated
RT   in mouse blastocysts during delayed implantation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RA   Hibbs M.L.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/C3; TISSUE=Liver;
RX   MEDLINE=94170986; PubMed=8125255;
RA   Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA   Jenkins N.A., Lee D.C.;
RT   "The mouse waved-2 phenotype results from a point mutation in the EGF
RT   receptor tyrosine kinase.";
RL   Genes Dev. 8:399-413(1994).
RN   [5]
RP   SEQUENCE OF 1-714 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=91232866; PubMed=2030916;
RA   Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT   "Comparison of EGF receptor sequences as a guide to study the ligand
RT   binding site.";
RL   Oncogene 6:673-676(1991).
RN   [6].
RP   SEQUENCE OF 969-1117 FROM N.A.
RC   STRAIN=C3H;
RA   Eisinger D.P., Serrero G.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X78987; CAA55587.1; -.
DR   EMBL; U03425; AAA17899.1; -.
DR   EMBL; X59698; CAA42219.1; -.
DR   EMBL; L06864; AAA53029.1; -.
DR   EMBL; Z12608; CAA78249.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1210       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       25    647       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    648    670       POTENTIAL.
FT   DOMAIN      671   1210       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       75    300       APPROXIMATE.
FT   REPEAT      390    600       APPROXIMATE.
FT   DOMAIN     1028   1071       SER-RICH.
FT   DOMAIN      714    981       PROTEIN KINASE.
FT   NP_BIND     720    728       ATP (BY SIMILARITY).
FT   BINDING     747    747       ATP (BY SIMILARITY).
FT   ACT_SITE    839    839       BY SIMILARITY.
FT   DISULFID    190    199       BY SIMILARITY.
FT   DISULFID    194    207       BY SIMILARITY.
FT   DISULFID    215    223       BY SIMILARITY.
FT   DISULFID    219    231       BY SIMILARITY.
FT   DISULFID    232    240       BY SIMILARITY.
FT   DISULFID    236    248       BY SIMILARITY.
FT   DISULFID    251    260       BY SIMILARITY.
FT   DISULFID    264    291       BY SIMILARITY.
FT   DISULFID    295    307       BY SIMILARITY.
FT   DISULFID    311    326       BY SIMILARITY.
FT   DISULFID    329    333       BY SIMILARITY.
FT   DISULFID    506    515       BY SIMILARITY.
FT   DISULFID    510    523       BY SIMILARITY.
FT   DISULFID    526    535       BY SIMILARITY.
FT   DISULFID    539    555       BY SIMILARITY.
FT   DISULFID    558    571       BY SIMILARITY.
FT   DISULFID    562    579       BY SIMILARITY.
FT   DISULFID    582    591       BY SIMILARITY.
FT   DISULFID    595    617       BY SIMILARITY.
```

```
FT   DISULFID    620    628       BY SIMILARITY.
FT   DISULFID    624    636       BY SIMILARITY.
FT   MOD_RES     680    680       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   MOD_RES    1092   1092       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1110   1110       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1172   1172       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1197   1197       PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT                                (BY SIMILARITY).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    175    175       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    352    352       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    444    444       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     19     19       C -> S (IN REF. 2).
FT   CONFLICT    539    539       C -> W (IN REF. 5).
FT   CONFLICT    991    991       L -> F (IN REF. 4).
FT   CONFLICT   1116   1117       HP -> DR (IN REF. 6).
SQ   SEQUENCE   1210 AA;  134853 MW;  690E20D46DF2D2F5 CRC64;

  Query Match             44.6%;  Score 3041;  DB 1;  Length 1210;
  Best Local Similarity   48.9%;  Pred. No. 1.6e-156;
  Matches  622;  Conservative  168;  Mismatches  369;  Indels  114;  Gaps   25;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
              ||  ||    :|| ||  :|    :  | |    |: :   |:|| ||||:||:  |
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
              ||||: |||| |||||| | | ::||: |:|:||  |:|: ||||:| |
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
                | :  ||||| :|:| ||| | |    || ||  ||| |:||    |  |  : |
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
                :  :|  | | |    |||    |:|| ||: :||  |: ||:|  |:|||| |||||
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              ||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:||  || ||
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              : || |||   |    ||  ||| ::|:||  || :|| |:|:   | ::  |  | ::
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
               |  |  | | |  || :| ||  :  | || | :|:: :|::||||:| | ||||:  ||
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
                |:||::||||   :| :|| : || |: |||||:|:  |  :|  | :||: :|: |
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
              :||  |:|      || | :|      |: ||:   |||| |  ||:|    ||:||||
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              :| :|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||  ||:
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
              |:  : : : :||: |    |  |  |||: |     :||        : || : :|| |
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 LVVVLGVVFGILIF---NNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILKETELR 719
```

```
              | :|: |  || :|    :     ||       :|||||||  |||| :|||||||  :
Db        657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFK 715

Qy        720 KVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPY 779
              |:|||||||||||||:|||:|| |||||||| ||| ||||||||||||||||  | :|:
Db        716 KIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPH 775

Qy        780 VSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRL 839
              | |||||||||||||:|||||||||||:|||::  :||| |||||:|||||:|||| ||
Db        776 VCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRL 835

Qy        840 VHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTH 899
              |||||||||||||:| |||||||||:||  :| ||||:|||||||||||||||  | :||
Db        836 VHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTH 895

Qy        900 QSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMI 959
              |||||||||||||||||:|||||||| :|  :|||||||||||||||||||||||||||
Db        896 QSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVKCWMI 955

Qy        960 DSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1018
              |:: ||:|||| |||:|||||||::||| :|  :    || || |||:|::::|| |:|||
Db        956 DADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVVDA 1015

Qy       1019 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1078
              :|||:|||||                | ||                  |:||  |
Db       1016 DEYLIPQQGFF----------------NSPST-----------------SRTPLLSSLS 1041

Qy       1079 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPE 1136
              | |:    :  :       |    : : ||||| |||  :  :   | ::      | ||
Db       1042 ATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------PVPE 1091

Qy       1137 YVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL- 1195
              |||| |  :|   : |:   :|          :|::   ::      || |||||
Db       1092 YVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNPEYLN 1140

Qy       1196 TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTP 1239
              | |          |   |  |::   | |       :|       |:   |   ||| |
Db       1141 TAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIFKG-P 1190

Qy       1240 TAENPEYLGLDVP 1252
              |||| ||| :  |
Db       1191 TAENAEYLRVAPP 1203


RESULT 6
ERB4_HUMAN
ID   ERB4_HUMAN     STANDARD;      PRT;  1308 AA.
AC   Q15303;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE   (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN   ERBB4 OR HER4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM JM-A).
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=93189574; PubMed=8383326;
RA   Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA   Foy L., Neubauer M.G., Shoyab M.;
RT   "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT   epidermal growth factor receptor family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97476287; PubMed=9334263;
RA   Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA   Klagsbrun M.;
```

```
RT   "A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
RT   tissue distribution and differential processing in response to
RT   phorbol ester.";
RL   J. Biol. Chem. 272:26761-26768(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC       FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC       BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC       NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC       CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC       KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC       PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC       LUNG, SALIVARY GLAND, AND PANCREAS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07868; AAB59446.1; -.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3432; ERBB4.
DR   MIM; 600543; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN      26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   652    675       POTENTIAL.
FT   DOMAIN     676   1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     186    334       CYS-RICH.
FT   DOMAIN     496    633       CYS-RICH.
FT   DOMAIN     718    985       PROTEIN KINASE.
FT   NP_BIND    724    732       ATP (BY SIMILARITY).
FT   BINDING    751    751       ATP (BY SIMILARITY).
FT   ACT_SITE   843    843       BY SIMILARITY.
FT   DISULFID   189    197       BY SIMILARITY.
FT   DISULFID   193    205       BY SIMILARITY.
FT   DISULFID   213    221       BY SIMILARITY.
FT   DISULFID   217    229       BY SIMILARITY.
FT   DISULFID   230    238       BY SIMILARITY.
FT   DISULFID   234    246       BY SIMILARITY.
FT   DISULFID   249    258       BY SIMILARITY.
FT   DISULFID   262    289       BY SIMILARITY.
FT   DISULFID   293    304       BY SIMILARITY.
FT   DISULFID   308    323       BY SIMILARITY.
FT   DISULFID   326    330       BY SIMILARITY.
FT   DISULFID   503    512       BY SIMILARITY.
FT   DISULFID   507    520       BY SIMILARITY.
FT   DISULFID   523    532       BY SIMILARITY.
FT   DISULFID   536    552       BY SIMILARITY.
FT   DISULFID   555    569       BY SIMILARITY.
FT   DISULFID   559    577       BY SIMILARITY.
FT   DISULFID   580    589       BY SIMILARITY.
FT   DISULFID   593    614       BY SIMILARITY.
FT   DISULFID   617    625       BY SIMILARITY.
FT   DISULFID   621    633       BY SIMILARITY.
FT   MOD_RES   1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD   138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   626    648       NGPTSHDCIYYPWTGHSTLPQHA -> IGSSIEDCIGLMD
FT                               (IN ISOFORM JM-B).
SQ   SEQUENCE   1308 AA;  146807 MW;  5E4AE80985D88761 CRC64;
```

```
  Query Match             42.7%;  Score 2908.5;  DB 1;  Length 1308;
  Best Local Similarity   44.4%;  Pred. No. 2.3e-149;
  Matches  600;  Conservative  182;  Mismatches  384;  Indels  185;  Gaps   28;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |      | || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  |  ||||: ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::||||| ||| : :|  ||| ||| |:||          |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||    || ||:   |  :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||   ||  |:| : |: |:  |  : | |:|   :  :        :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :    : ||:| || |: ||||:| | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||   :| ::   ||  : |  ||: |: ||:   |||||| ||::| :| ||: |:|
```

```
Db       476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy       544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602

Db       536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy       603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PAEQRASPL 651

Db       596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQHAR-TPL 652

Qy       652 TSIVSAVV-GILLVVVLGVVFGILIFNNFTVSFWLRVPKVSASH-------LEPLTPSGA 703

Db       653 --IAAGVIGGLFILVIVGLTFAVYVRRK-------SIKKKRALRRFLETELVEPLTPSGT 703

Qy       704 MPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 763

Db       704 APNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKANV 763

Qy       764 EILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNW 823

Db       764 EFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLLNW 823

Qy       824 CMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVP 883

Db       824 CVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKMP 883

Qy       884 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPP 943

Db       884 IKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQPP 943

Qy       944 ICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDSTF 1002

Db       944 ICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDSKF 1003

Qy      1003 YRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLE 1062

Db      1004 FQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIGHS 1051

Qy      1063 PSEEEAPRS----------------------------PLAP-SEGAGSDVFDGDLGMGAA 1093

Db      1052 PPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNGTL 1111

Qy      1094 KGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQ 1146

Db      1112 RKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE---- 1167

Qy      1147 PPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPH 1206

Db      1168 -----ENPFVSRR---------------KNGDLQ-------ALDNPEYHNASNG------ 1194

Qy      1207 PPPA---------------------------------FSPAFDNLYYWDQDPPERGA-- 1230

Db      1195 PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRSTLQ 1254

Qy      1231 PPSTFKGTPT-------------AENPEYL 1247

Db      1255 HPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 7
ERB4_RAT
ID   ERB4_RAT       STANDARD;      PRT;  1308 AA.
AC   Q62956; Q9Z2N7;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN   ERBB4 OR TYRO-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=98221155; PubMed=9553078;
RA   Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA   Marchionni M.A., Kelly R.A.;
RT   "Neuregulins promote survival and growth of cardiac myocytes.
RT   Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT   ventricular myocytes.";
RL   J. Biol. Chem. 273:10261-10269(1998).
RN   [2]
RP   SEQUENCE OF 848-901 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   SEQUENCE OF 1031-1198 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC       NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC       OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC       RETICULAR NUCLEUS OF THE THALLAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC       HEART.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF041838; AAD08899.1; -.
DR   EMBL; U52531; AAC53051.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL       1     25       POTENTIAL.
```

```
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN       676   1308      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       186    334      CYS-RICH.
FT   DOMAIN       496    633      CYS-RICH.
FT   DOMAIN       718    985      PROTEIN KINASE.
FT   NP_BIND      724    732      ATP (BY SIMILARITY).
FT   BINDING      751    751      ATP (BY SIMILARITY).
FT   ACT_SITE     843    843      BY SIMILARITY.
FT   DISULFID     189    197      BY SIMILARITY.
FT   DISULFID     193    205      BY SIMILARITY.
FT   DISULFID     213    221      BY SIMILARITY.
FT   DISULFID     217    229      BY SIMILARITY.
FT   DISULFID     230    238      BY SIMILARITY.
FT   DISULFID     234    246      BY SIMILARITY.
FT   DISULFID     249    258      BY SIMILARITY.
FT   DISULFID     262    289      BY SIMILARITY.
FT   DISULFID     293    304      BY SIMILARITY.
FT   DISULFID     308    323      BY SIMILARITY.
FT   DISULFID     326    330      BY SIMILARITY.
FT   DISULFID     503    512      BY SIMILARITY.
FT   DISULFID     507    520      BY SIMILARITY.
FT   DISULFID     523    532      BY SIMILARITY.
FT   DISULFID     536    552      BY SIMILARITY.
FT   DISULFID     555    569      BY SIMILARITY.
FT   DISULFID     559    577      BY SIMILARITY.
FT   DISULFID     580    589      BY SIMILARITY.
FT   DISULFID     593    614      BY SIMILARITY.
FT   DISULFID     617    625      BY SIMILARITY.
FT   DISULFID     621    633      BY SIMILARITY.
FT   MOD_RES     1162   1162      PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     1188   1188      PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     1258   1258      PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     1284   1284      PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     138    138      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     174    174      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     253    253      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     358    358      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     410    410      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     473    473      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     495    495      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     548    548      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     620    620      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    1062   1062      S -> N (IN REF. 3).
FT   CONFLICT    1080   1082      PYT -> SYR (IN REF. 3).
SQ   SEQUENCE    1308 AA;  146957 MW;  D944BB0996A08B41 CRC64;

  Query Match             42.4%;  Score 2889;  DB 1;  Length 1308;
  Best Local Similarity   44.2%;  Pred. No. 2.6e-148;
  Matches  598;  Conservative  189;  Mismatches  393;  Indels  172;  Gaps   28;

Qy          1 MELA-ALCRWGLLL--ALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQ 57
              |:||  |  || ||  |     :||  || ||: ||   :  |    ||  |: |:||
Db          1 MKLATGLWVWGSLLVAARTVQPSASQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVM 60

Qy         58 GNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVL 117
              ||||:| :  |  ||||: |:|| ||||:| || | :||: |||:|||:|:|| ||||:
Db         61 GNLEITSIEHNRDLSFLRSIREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIF 120

Qy        118 DNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNN 177
               |      |          ||:|| |::||||| ||| : :|  ||| ||| |:||
Db        121 LNYRKDGNF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPW 171

Qy        178 QLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCC 236
              :||: |  |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||
Db        172 PSNMTLVSTIGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCC 230

Qy        237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
              | :|| ||:||| :|| ||::|| || |   ||    ||  ||:   |  :||:|| ||
Db        231 HRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCV 290

Qy        297 TACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF 356
              ||:|::  |  ||   ||     ||  |:| : |: |:  | : | |:|   : :
Db        291 KKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTV 348

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
                :|  | || |:| ||      ||| :    : ||:| || |: |||||:| |  ||
Db        349 DSSNIDKFINCTKINGNLIFLVTGIHGDPYNAIDAIDPEKLNVFRTVREITGFLNIQTWP 408

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
               :: | ||| ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||:
Db        409 PNMTDFSVFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCY 468

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
              ||: |  ||   :| ::   ||  : |  ||: |: ||:   ||||||  ||::| :| |
Db        469 YHTINWTTLFSTVNQRIVIRDNRRAENCTAEGMVCNHLCSNDGCWGPGPDQCLSCRRFSR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPF 595
              |: |:|  | :  |  ||: |   |: |  :|:  ::| :|| ||  | |  |:|:|| |
Db        529 GKICIESCNLYDGEFREFENGSICVECDSQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPN 588

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PA 644
              || :||  ::   |:  |:|: |::  | ||  |||  |        |          |
Db        589 CVEKCPDVLQGANSF--IFKYADQDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQ 646

Qy        645 EQRASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTVSFWLRVPKVSASH-------LE 696
              | :||  | : |: |: ::|:: : | : :           : | |          :|
Db        647 HAR-TPL--IAAGVIGGLFILVIMALTFAVYVRRK-------SIKKKRALRRFLETELVE 696

Qy        697 PLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLREN 756
              ||||||  |||||:|||||||::|||||||||||||||||:|:|| |||||||:| |
Db        697 PLTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNET 756

Qy        757 TSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLG 816
              | |||| | :||| :|| |  |:: ||||:||: |:|||||||:|||||::| |::  :|
Db        757 TGPKANVEFMDEALIMASVDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIG 816

Qy        817 SQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYH 876
              || |||||:|||||| ||||: |||||||||||||||||||||||||||||||: || ||:
Db        817 SQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYN 876

Qy        877 ADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKG 936
              |||||:|||||||| |  |:|||||||||||:|||||||  |||||||  ||||||||||
Db        877 ADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKG 936

Qy        937 ERLPQPPICTIDVYIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPA 995
              |||||||||||||::||||||:: ||:|:|| :|||||||||||::||| :| :
Db        937 ERLPQPPICTIDVYIVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLP 996

Qy        996 SPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDP---------------------- 1033
              || || |:::||:::|: |::|||||||| | |  | |
Db        997 SPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTSRTRIDSNRSEIGHSPPPA 1055

Qy       1034 -APGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGA 1092
               | :|    ::    :|: |   : :    :     | :|:|  :|| ::::||     |
Db       1056 YTPMSGSQFVYQDGGFATQQG---MPMPYTATTSTIPEAPVA--QGATAEMFDDSCCNGT 1110

Qy       1093 AKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRP 1145
                :       : |  |||| ||||  |        : :||: |:   |: ||:|  :
Db       1111 LRKPVVPHVQEDSSTQRYSADPTVFAPERNPRAELDEEGYMTPMHDKPKQEYLNPVE--- 1167

Qy       1146 QPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQP 1205
                | |  : |                ||| ::       |::|||| :  |
Db       1168 ------ENPFVSRR--------------KNGDLQ-------ALDNPEYHSASSG----- 1194

Qy       1206 HPPPA---------------------------------FSPAFDNLYYWDQDPPERGA- 1230
                ||  |                                 ||||  ||:   | |
Db       1195 -PPKAEDEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPDYWNHSLPPRSTL 1253

Qy       1231 -PPSTFKGTPT--------------AENPEYL 1247
                |   :  |                |||||||
Db       1254 QHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285
```

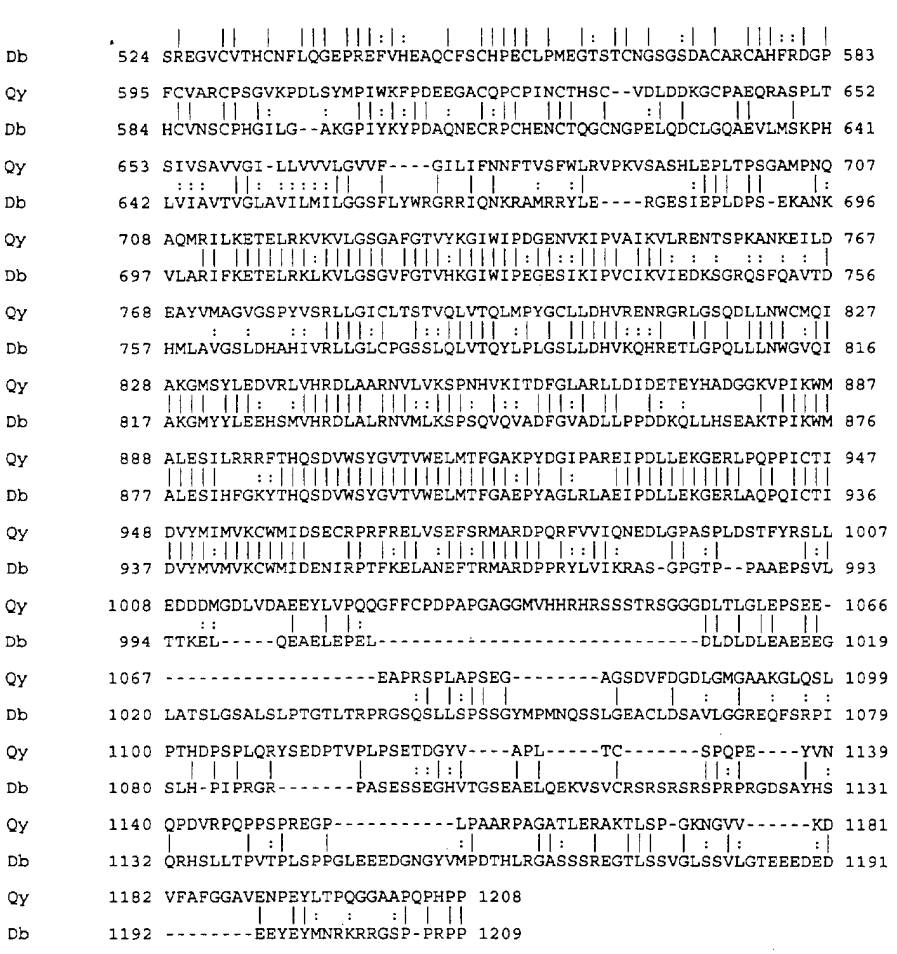
```
RESULT 8
XMRK_XIPMA
ID   XMRK_XIPMA      STANDARD;      PRT;  1167 AA.
AC   P13388;
DT   01-JAN-1990 (Rel. 13, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN   XMRK OR TU.
OS   Xiphophorus maculatus (Southern platyfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8083;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90015140; PubMed=2797166;
RA   Wittbrodt J., Adam D., Malitschek B., Maueler W., Raulf F.,
RA   Telling A., Robertson S.M., Schartl M.;
RT   "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT   inducing Tu locus in Xiphophorus.";
RL   Nature 341:415-421(1989).
RN   [2]
RP   REVISION TO 515.
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16891; CAA34770.2; -.
DR   PIR; S06142; S06142.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT   SIGNAL        1     25
FT   CHAIN        26   1167       MELANOMA RECEPTOR PROTEIN-TYROSINE
FT                                KINASE.
FT   DOMAIN        26    642       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     643    665       POTENTIAL.
FT   DOMAIN       666   1167       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       710    977       PROTEIN KINASE.
FT   NP_BIND      716    724       ATP (BY SIMILARITY).
FT   BINDING      743    743       ATP (BY SIMILARITY).
FT   ACT_SITE     835    835       BY SIMILARITY.
FT   DISULFID     195    204       BY SIMILARITY.
FT   DISULFID     199    212       BY SIMILARITY.
FT   DISULFID     220    228       BY SIMILARITY.
FT   DISULFID     224    236       BY SIMILARITY.
FT   DISULFID     237    245       BY SIMILARITY.
FT   DISULFID     241    253       BY SIMILARITY.
FT   DISULFID     256    265       BY SIMILARITY.
FT   DISULFID     269    296       BY SIMILARITY.
FT   DISULFID     300    311       BY SIMILARITY.
FT   DISULFID     315    330       BY SIMILARITY.
FT   DISULFID     333    337       BY SIMILARITY.
FT   DISULFID     504    513       BY SIMILARITY.
FT   DISULFID     508    521       BY SIMILARITY.
FT   DISULFID     524    533       BY SIMILARITY.
FT   DISULFID     537    553       BY SIMILARITY.
FT   DISULFID     556    569       BY SIMILARITY.
FT   DISULFID     560    577       BY SIMILARITY.
FT   DISULFID     593    615       BY SIMILARITY.
FT   DISULFID     618    626       BY SIMILARITY.
FT   DISULFID     622    634       BY SIMILARITY.
FT   CARBOHYD     114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1167 AA;  129934 MW;  4793E0749DC1D55A CRC64;

  Query Match             38.4%;  Score 2614.5;  DB 1;  Length 1167;
  Best Local Similarity   44.2%;  Pred. No. 1.4e-133;
  Matches  562;  Conservative  168;  Mismatches  394;  Indels  147;  Gaps   30;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              ||| :  ||| |       ||    :|| ||  :: :    : |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||   |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: | |: |
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 | ::|       ||::||| :||||| ||| :  || ||  :|| | ||  | :
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
               : ||      | |  |    |      ||      ||  |: :||  |  ||:|| | |||:|
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
               || ||||: :|||||  ||  | |:  ||     |:  : : : ||  :|||||:||
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
              || ||: |: |:|   |    ||  |:| : |: |   | :|| |:|:  : :|:  :
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
               | :  |: | || | :     ||:|||      : || |    |::|||||  | ||:
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
              ::  |||||||::||||   :  :| : :|   : |||||||:|: :|  :: :   | :
Db        417 NMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRY 476

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
               :|: | :|||:  |:: : |          |   |:  |:  |||||| ||:|    |
Db        477 ANTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCVSCLHVDR 529

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFC 596
```

```
Db        530 GGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKSAHFQDGPQC 589

Qy        597 VARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVS 656

Db        590 IPRCPHGILGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGD-IVSHSSLAVG 647

Qy        657 AVVGILLVVVLGVVFGIL-----IFNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMR 711

Db        648 LVSGLLITVIVALLIVVLLRRRRIKRKRTIRCLLQEKEL----VEPLTPSGQAPNQAFLR 703

Qy        712 ILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 771

Db        704 ILKETEFKKDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQEVLDEAYV 763

Qy        772 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 831

Db        764 MASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCVQIAKGM 823

Qy        832 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 891

Db        824 NYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPIKWMALES 883

Qy        892 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 951

Db        884 ILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPICTIEVYM 943

Qy        952 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1011

Db        944 IILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSLSDRRLFSRLLSSDD 1000

Qy       1012 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1071

Db       1001 --DVVDADEYLLPYKRI---------------------NRQGS----------------E 1021

Qy       1072 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDGYVAPLT 1130

Db       1022 PCIPPTGH--------------------PVRENSITLRNISDPTQNALEKDLDGH----- 1056

Qy       1131 CSPQPEYVNQPDVRPQP--------PSPRE-----GPLP-AARPAGATLERAKTLSPGKNG 1177

Db       1057 -----EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYLNTNQNS 1111

Qy       1178 VVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKG 1237

Db       1112 L---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGALTGNGMF 1146

Qy       1238 TPTAENPEYLG 1248

Db       1147 LPAAENLEYLG 1157
```

```
RESULT 9
ERB3_HUMAN
ID   ERB3_HUMAN     STANDARD;      PRT;  1342 AA.
AC   P21860;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN   ERBB3 OR HER3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90083234; PubMed=2687875;
RA   Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT   "Isolation and characterization of ERBB3, a third member of the
RT   ERBB/epidermal growth factor receptor family: evidence for
RT   overexpression in a subset of human mammary tumors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90311312; PubMed=2164210;
RA   Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA   Todaro G.J., Shoyab M.;
RT   "Molecular cloning and expression of an additional epidermal growth
RT   factor receptor-related gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN   [3]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RC   TISSUE=Placenta;
RX   MEDLINE=93282822; PubMed=7685162;
RA   Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT   "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT   tyrosine kinase.";
RL   Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC       SECRETED (SHORT FORM).
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC       SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC   -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29366; AAA35790.1; -.
DR   EMBL; M34309; AAA35979.1; -.
DR   EMBL; S61953; AAB26935.1; -.
DR   PIR; A36223; A36223.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3431; ERBB3.
DR   MIM; 190151; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1342       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN      20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   644    664       POTENTIAL.
FT   DOMAIN     665   1342       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     709    966       PROTEIN KINASE.
FT   NP_BIND    715    723       ATP (BY SIMILARITY).
```

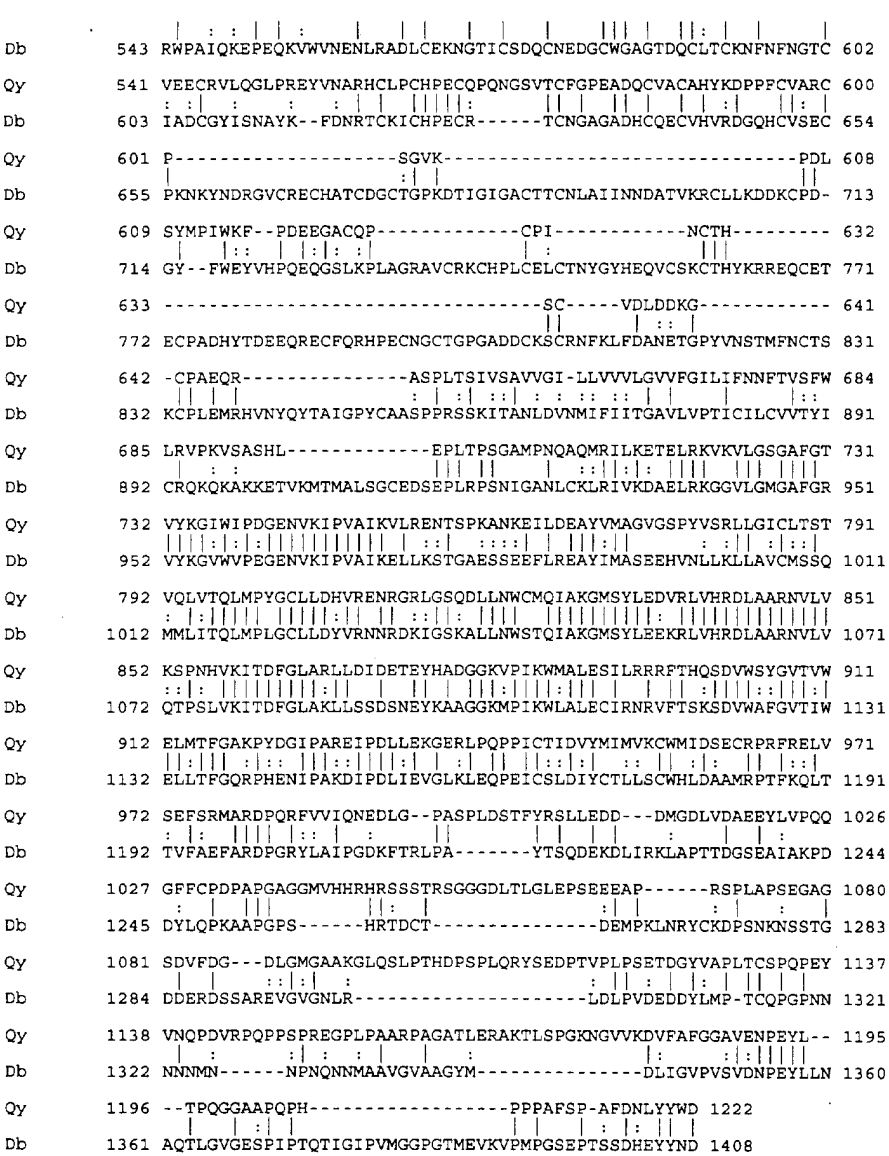
```
FT   BINDING     742    742       ATP (BY SIMILARITY).
FT   ACT_SITE    834    834       BY SIMILARITY.
FT   DISULFID    186    194       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    552    565       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    141    183       EILSGGVYIEKNDKLCHMDTIDWRDIVRDRDAEIVVKDNGR
FT                                SC -> GQFPMVPSGLTPQPAQDWYLLDDDPRLLTLSASSK
FT                                VPVTLAAV (IN SHORT ISOFORM).
FT   VARSPLIC    184   1342       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    560    560       E -> G (IN REF. 2).
FT   CONFLICT   1064   1064       E -> G (IN REF. 2).
SQ   SEQUENCE   1342 AA;  148097 MW;  7201E7F66CA374BD CRC64;

  Query Match             35.1%;  Score 2389.5;  DB 1;  Length 1342;
  Best Local Similarity   40.0%;  Pred. No. 2.2e-121;
  Matches  527;  Conservative  195;  Mismatches  453;  Indels  141;  Gaps   35;

Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
              |||  :|          | :|| ||     :          |  ||: |:|| ||||:
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
              || ||||| |:|| ||||:| |:    :|| ||:|||||:::  :|: |:     ||  |
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
                   |   ||:|:|   |||||  |||  |::| :||: ||| |:||       :: :
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
                |:| ||  :|||  ||||   ||||||:|:|:|:|   |   |   |||::||  ||:|
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
              |: :|| || |||  || |  |     ||   :  ||   ||:  |||  :|  || :||:|::
Db        238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFQLEPNPHTKYQYGGVCVASCPHNFV-V 296

Qy        307 DVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF--IGITELE- 363
              |   ||   ||    ||   ::|: ||   |  |  | |     : |:|   :   :  ::
Db        297 DQTSCVRACPPDKMEVD-KNGLKMCEPCGGLCPKACEGTG-----SGSRFQTVDSSNIDG 350

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |   ||  |:  |:| ||           |  ||:| ||  |: ||||||  |  :  :  |
Db        351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVFRTVREITGYLNIQSWPPHMHNFS 410

Qy        424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
              || ||  | || |:|  :| | ::  | :: || |||:|: :|   |   |  ||: |:: |
Db        411 VFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNW 470

Qy        483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECV 541
              :: |  | :  |   |||  :|| ||  |   ||: | |||||| ||::|  :  || ||
Db        471 TKVLRGPTEERLDIKHNRPRRDCVAEGKVCDPLCSSGGCWGPGPGQCLSCRNYSRGGVCV 530

Qy        542 EECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCP 601
                |  |  |||: :    |  ||||||   |: ||  |  :| |   |||::| | ||: ||
Db        531 THCNFLNGEPREFAHEAECFSCHPECQPMEGTATCNGSGSDTCAQCAHFRDGPHCVSSCP 590

Qy        602 SGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRA----SPLTSIVSA 657
               ||      :  ||:|:|| :  |:|| ||| |   : : |  :         : || ::
Db        591 HGVLG--AKGPIYKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLIGKTHLTMALTV 648

Qy        658 VVGILLV-VVLGVVF----GILIFNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRI 712
              : |:::: ::|| |     |  |  |      :  :         :||| ||  |:  ||
Db        649 IAGLVVIFMMLGGTFLYWRGRRIQNKRAMRRYLE----RGESIEPLDPS-EKANKVLARI 703

Qy        713 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 772
               |||||||:|||||  ||||:||:||:||:|::||||  |||: :  : : :   :  :
Db        704 FKETELRKLKVLGSGVFGTVHKGVWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAI 763

Qy        773 AGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMS 832
                :   :: |||:|  |::||||| :| |  ||||||:::|| ||  |  ||||  :||||||
Db        764 GSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVRQHRGALGPQLLLNWGVQIAKGMY 823

Qy        833 YLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESI 892
              |||:  :|||:|||||||:|||:  |::  |||:| ||  |: :        | |||||||||
Db        824 YLEEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKWMALESI 883

Qy        893 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 952
                ::|:|||||||||||||||||||||| :| |   |  ||||||||||  |  ||||||:
Db        884 HFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEVPDLLEKGERLAQPQICTIDVYMV 943

Qy        953 MVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPA---SPLDSTFYRSLLED 1009
              |||||||   || |:|| :|:||||||  |::||: |   ||             ||:
Db        944 MVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKLEE 1002

Qy       1010 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP-SEEEA 1068
               ::    :| :   |  ::                    : :| :  |  |:| :   :
Db       1003 VELEPELDLDLDLEAEED------------------NLATTTLGSALSLPVGTLNRPRG 1043

Qy       1069 PRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHD-PSPLQRYSEDPTVPLP------SE 1121
               :| |:||  |    : |:||    :    |    |   ||  |: :       |:|
Db       1044 SQSLLSPSSGY-MPMNQGNLGESCQESAVSGSSERCPRPVSLH------PMPRGCLASES 1096

Qy       1122 TDGYVA---------------PLTCSPQPE----YVNQPDVRPQPPSPREGP------- 1154
              ::|:|                |:|:|   |  |  |    |   |:|  |
Db       1097 SEGHVTGSEAELQEKVSMCRSRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPPGLEEEDV 1156

Qy       1155 ----LPAARPAGATLERAKTLSP-GKNGVV------KDVFAFGGAVENPEYLTPQGGAAP 1203
                  :|   |    |  |||  |  : |:        :|       |  ||:   :   :|
Db       1157 NGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEEEDED--------EEYEYMNRRRRHSP 1208

Qy       1204 QPHPPPAFSPAFDNLYYWD-----------QDPPERGAPPSTFKGTPTAENPEYL 1247
               ||||    |         | |            |    |     |  ||  |: ||:
Db       1209 -PHPPRPSSLEELGYEYMDVGSDLSASLGSTQSCPLHPVPIMPTAGTTPDEDYEYM 1263

RESULT 10
ERB3_RAT
ID   ERB3_RAT       STANDARD;      PRT;  1339 AA.
AC   Q62799; Q62955;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3).
GN   ERBB3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96096535; PubMed=8522190;
RA   Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT   "Cloning of the rat ErbB3 cDNA and characterization of the
RT   recombinant protein.";
RL   Gene 165:279-284(1995).
RN   [2]
RP   REVISIONS TO 85; 513 AND 565.
RA   Hellyer N.J., Koland J.G.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 922-1097 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29339; AAC28498.2; -.
DR   EMBL; U52530; AAC53050.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1339       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN       20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    644    662       POTENTIAL.
FT   DOMAIN      663   1339       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      183    259       CYS-RICH.
FT   DOMAIN      707    964       PROTEIN KINASE.
FT   NP_BIND     713    721       ATP (BY SIMILARITY).
FT   BINDING     740    740       ATP (BY SIMILARITY).
FT   ACT_SITE    832    832       BY SIMILARITY.
FT   DISULFID    186    194       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1028   1028       L -> P (IN REF. 3).
SQ   SEQUENCE   1339 AA;  147545 MW;  0AA5F2402BBFDF1E CRC64;

  Query Match           34.0%;  Score 2315.5;  DB 1;  Length 1339;
  Best Local Similarity  40.4%;  Pred. No. 2.1e-117;
  Matches  520;  Conservative  171;  Mismatches  431;  Indels  165;  Gaps   36;

Qy          3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              |  ||     |   |  |:    |  || ||   | :    :      |  ||: |:|| ||
Db          7 LQVLC----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLYKLYEKCEVVMGN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:     || ||||| |:|| ||||:| |:   :||  ||:|||||::  :|: |:
Db         63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLPNLRVVRGTQVYDGKFAIFVM-- 120

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 ||  |      |   ||:|:   ||||| ||| |::| :||: ||| |:||
Db        121 ---LNYNT----NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRVR--- 170

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
                ::  |    | ||  :||| ||||   :||| ||:|:||  |  || || |  |||:
Db        171 GAEIVVKNNGANCPPCHEVCKG-RCWGPGPDDCQILTKTICAPQCNGRCFGPNPNQCCHD 229

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              :|| ||:||: :|| ||  || || |   ||  : ||  ||:  |||  :| :|  || :
Db        230 ECAGGCSGPQDTDCFACRRFNDSGACVPRCPEPLVYNKLTFQLEPNPHTKYQYGGVCVAS 289

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGL--GMQYIKANSKF 356
              ||:|::  |   |   ||  ||  : | : || |   | : | |   | :|   :|
Db        290 CPHNFV-VDQTFCVRACPPDKMEVD-KHGLKMCEPCGGLCPKACEGTGSGSRYQTVDSSN 347

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
              |      |  | || |:| ||    : ||       | ||:| || |: ||||| | :||
Db        348 ID----GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLNVFRTVREITGYLNIQSWP 403

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLC 475
              : : ||| ||  | || |:|  :| | :: | :: || |||:|: :|   |  | ||
Db        404 PHMHNFSVFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLC 463

Qy        476 FVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQF 534
              : |:: | :| | | :  |    :||  ||: ||  |  ||: | |||||| ||::|  :
Db        464 YHHSLNWTRLLRGPSEERLDIKYDRPLGECLAEGKVCDPLCSSGGCWGPGPGQCLSCRNY 523

Qy        535 LRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPP 594
```

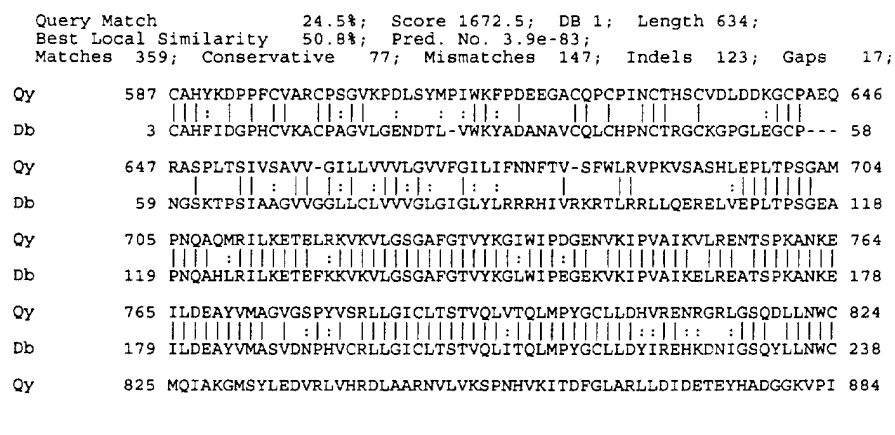


```
RESULT 11
EGFR_DROME
ID   EGFR_DROME     STANDARD;      PRT;  1426 AA.
AC   P04412; O61601; Q9W2G0; P81868;
DT   13-AUG-1987 (Rel. 05, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE   (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN   EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX   MEDLINE=94350209; PubMed=8070664;
RA   Clifford R., Schupbach T.;
RT   "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT   that several genetically defined classes of alleles cluster in
RT   subdomains of the receptor protein.";
RL   Genetics 137:531-550(1994).
RN   [2]
RP   REVISIONS.
RA   Clifford R., Schupbach T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=85124611; PubMed=2982499;
RA   Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT   "The Drosophila EGF receptor gene homolog: conservation of both
RT   hormone binding and kinase domains.";
RL   Cell 40:599-607(1985).
RN   [4]
RP   SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC   STRAIN=Oregon-R; TISSUE=Embryo;
RX   MEDLINE=87002474; PubMed=3093080;
RA   Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT   "Alternative 5' exons and tissue-specific expression of the
RT   Drosophila EGF receptor homolog transcripts.";
RL   Cell 46:1091-1101(1986).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=99102120; PubMed=9882502;
RA   Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT   "Several levels of EGF receptor signaling during photoreceptor
RT   specification in wild-type, Ellipse, and null mutant Drosophila.";
RL   Dev. Biol. 205:129-144(1999).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [7]
RP   SEQUENCE OF 959-1078 FROM N.A.
```

```
RC   STRAIN=Daekwanryeong;
RX   MEDLINE=85137938; PubMed=2983232;
RA   Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT   "A Drosophila genomic sequence with homology to human epidermal
RT   growth factor receptor.";
RL   Nature 314:178-180(1985).
RN   [8]
RP   SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=92038942; PubMed=1936959;
RA   Raz E., Schejter E.D., Shilo B.Z.;
RT   "Interallelic complementation among DER/flb alleles: implications for
RT   the mechanism of signal transduction by receptor-tyrosine kinases.";
RL   Genetics 129:191-201(1991).
RN   [9]
RP   REVIEW.
RX   MEDLINE=97248481; PubMed=9094709;
RA   Perrimon N., Perkins L.A.;
RT   "There must be 50 ways to rule the signal: the case of the Drosophila
RT   EGF receptor.";
RL   Cell 89:13-16(1997).
CC   -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC       WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC       MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC       CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC       DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC       POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC       ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC       AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC       SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC       CUTICLE.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC       MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC       PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC       TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC       UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC       TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC       EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC       IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC       POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC       AND THORACIC AND ABDOMINAL GANGLIA.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF052754; AAC08536.1; -.
DR   EMBL; AF052753; AAC08536.1; JOINED.
DR   EMBL; AF052754; AAC08535.1; -.
DR   EMBL; AF052752; AAC08535.1; JOINED.
DR   EMBL; K03054; AAA51462.1; -.
DR   EMBL; K03417; AAA51460.1; -.
DR   EMBL; K03416; AAA50965.1; -.
DR   EMBL; K03418; AAA51461.1; -.
DR   EMBL; AF109077; AAD26134.1; -.
DR   EMBL; AF109078; AAD26132.1; -.
DR   EMBL; AF109082; AAD26132.1; JOINED.
DR   EMBL; AF109078; AAD26133.1; -.
DR   EMBL; AF109084; AAD26133.1; JOINED.
DR   EMBL; AF109079; AAD26130.1; -.
DR   EMBL; AF109081; AAD26130.1; JOINED.
DR   EMBL; AF109079; AAD26131.1; -.
DR   EMBL; AF109083; AAD26131.1; JOINED.
DR   EMBL; AF109080; AAD26135.1; -.
DR   EMBL; AE003454; AAF46732.1; -.
DR   EMBL; X02293; CAA26157.1; -.
DR   EMBL; X78920; CAA55523.1; -.
DR   EMBL; X78918; CAA55521.1; -.
DR   EMBL; X78919; CAA55522.1; -.
DR   PIR; A00640; GQFFE.
DR   HSSP; P11362; 1FGK.
DR   FlyBase; FBgn0003731; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 7.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW   Developmental protein.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1426       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    868       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    869    889       POTENTIAL.
FT   DOMAIN      890   1426       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      938   1198       PROTEIN KINASE.
FT   NP_BIND     944    952       ATP (BY SIMILARITY).
FT   BINDING     971    971       ATP (BY SIMILARITY).
FT   ACT_SITE   1063   1063       BY SIMILARITY.
FT   MOD_RES     902    902       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

  Query Match             28.2%;  Score 1924;  DB 1;  Length 1426;
  Best Local Similarity   32.1%;  Pred. No. 2.7e-96;
  Matches  458;  Conservative  189;  Mismatches  433;  Indels  348;  Gaps   41;

Qy         24 QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT-NASLSFLQDIQEVQG 82
              ::| || :| :|:: | | || | | |||:||:|| | |||| :|:|| |
Db        100 KICIGTKSRLSVPSNKEHHYRNLRDRYTNCTYVDGNLKLTWLPNENLDLSFLDNIREVTG 159

Qy         83 YVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGL 137
              :|||:| |::|   :|:|:|| ||     |: ||| |              : :
Db        160 YILISHVDVKKVVFPKLQIIRGRTLFSLSVEEEKYALFV----------------TYSKM 203

Qy        138 RELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSP 197
                |::  | ::| | |    | ||: || | :|           |  | |  |
Db        204 YTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPKCHE 263

Qy        198 MCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDCLAC 255
                |  |||| ::|| ::  |:  ||  || || | :||| || ||||  ||:||
Db        264 SCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDCIAC 322

Qy        256 LHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVC 315
              :|   : :  || :  ||  |:   ||||:| :||:|| || :|  | |:|  |
Db        323 KNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACVRSC 381

Qy        316 PLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAGCKK 369
              |   :   |    |  |: || : |                |:| |      | |
Db        382 PQDKMDKGGE-----CVPCNGPCPKTC-------------PGVTVLHAGNIDSFRNCTV 422

Qy        370 IFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDSLPD 421
              | |::  | ::| |  |  :|        || ||: :|| |::|||||| |      :
Db        423 IDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQFRN 482

Qy        422 LSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              || :|||: | || |    : :| :    :  | :|:|::: ||  :| ||  ||:|  :
Db        483 LSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYVSNI 542

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
```



```
RESULT 12
ERBB_ALV
ID   ERBB_ALV       STANDARD;      PRT;   634 AA.
AC   P00534;
DT   21-JUL-1986 (Rel. 01, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian leukosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11864;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85228222; PubMed=2988784;
RA   Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA   Crittenden L.B., Raines M.A., Kung H.-J.;
RT   "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT   processing and promoter insertion result in expression of an
RT   amino-truncated EGF receptor.";
RL   Cell 41:719-726(1985).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC       IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC       AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC       THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC       C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC       PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10066; AAA48763.1; ALT_INIT.
DR   PIR; A00643; TVCHLV.
DR   PIR; B00643; TVFVLV.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
SQ   SEQUENCE   634 AA;  70891 MW;  E705E33A0BE01FCC CRC64;

  Query Match             24.5%;  Score 1672.5;  DB 1;  Length 634;
  Best Local Similarity   50.8%;  Pred. No. 3.9e-83;
  Matches  359;  Conservative   77;  Mismatches  147;  Indels  123;  Gaps   17;
```



```
Db        239 VQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPI 298

Qy        885 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 944

Db        299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy        945 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1003

Db        359 CTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy       1004 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1063

Db        419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------ 449

Qy       1064 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1118

Db        450 -----SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNF 496

Qy       1119 PSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKN 1176

Db        497 LEESIDDGFL------PAPEYVNQ--LMPKKPS--------------------TAMVQN 527

Qy       1177 GVVKDVF---------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYW 1221

Db        528 QIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFESSPYW 579

Qy       1222 DQ--------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1252

Db        580 IQSGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 625
```

```
RESULT 13
ERBB_AVIER
ID   ERBB_AVIER     STANDARD;      PRT;   604 AA.
AC   P00535;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ES4).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=79685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H;
RX   MEDLINE=84026539; PubMed=6313229;
RA   Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT   "The erbB gene of avian erythroblastosis virus is a member of the src
RT   gene family.";
RL   Cell 35:71-78(1983).
RN   [2]
RP   SEQUENCE OF 1-152 FROM N.A.
RX   MEDLINE=84223957; PubMed=6328658;
RA   Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA   Saule S., Martin P., Stehelin D.;
RT   "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT   new type of oncogene.";
RL   Science 224:1456-1459(1984).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC       ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC       IN CHICKENS.
CC   -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC       RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02006; AAA42394.1; ALT_INIT.
DR   EMBL; K01216; AAA42400.1; -.
DR   PIR; A00644; TVYUH.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132     399       PROTEIN KINASE.
FT   NP_BIND     138     146       ATP (BY SIMILARITY).
FT   BINDING     165     165       ATP (BY SIMILARITY).
FT   ACT_SITE    257     257       BY SIMILARITY.
FT   CONFLICT     29      29       R -> W (IN REF. 2).
FT   CONFLICT    140     140       S -> F (IN REF. 2).
FT   CONFLICT    146     146       I -> V (IN REF. 2).
SQ   SEQUENCE   604 AA;  67633 MW;  76EBCDD06745D609 CRC64;
```

Query Match 23.9%; Score 1626; DB 1; Length 604;
Best Local Similarity 50.7%; Pred. No. 1.2e-80;
Matches 349; Conservative 74; Mismatches 140; Indels 126; Gaps 16;

```
Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646

Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLTPSGAM 704

Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEA 118

Qy        705 PNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 764

Db        119 PNQAHLRILKETEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKELREATSPKANKE 178

Qy        765 ILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWC 824

Db        179 ILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWC 238

Qy        825 MQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPI 884

Db        239 VQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPI 298

Qy        885 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 944

Db        299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy        945 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1003

Db        359 CTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy       1004 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1063

Db        419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------ 449

Qy       1064 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1118

Db        450 -----SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNF 496

Qy       1119 PSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKN 1176

Db        497 LEESIDDGFL------PAPEYVNQ--LMPKKPSTAM------------------------ 524

Qy       1177 GVVKDVFAF------------------GGAVENPEYLTPQGGAAPQPHPPPAFSPAFDN 1217
```

```
Db       525 -VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFES 575

Qy      1218 LYYWDQDPPERGAPPSTFKGTPTAENPEY 1246
Db       576 SPYWIQSGNHQ----------INLDNPDY 594


RESULT 14
ERBB_AVIEU
ID   ERBB_AVIEU     STANDARD;      PRT;   540 AA.
AC   P11273;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ts167).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=103898;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87064458; PubMed=2878364;
RA   Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT   "A single amino acid substitution in v-erbB confers a thermolabile
RT   phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT   cells.";
RL   Mol. Cell. Biol. 6:1751-1759(1986).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13179; AAA42401.1; -.
DR   PIR; A25231; TVFVEB.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   VARIANT     270    270       H -> D (IN THERMOLABILE V-ERBB).
SQ   SEQUENCE   540 AA;  60412 MW;  5B53297AA068B65D CRC64;

  Query Match             22.8%;  Score 1553;  DB 1;  Length 540;
  Best Local Similarity   53.4%;  Pred. No. 8.8e-77;
  Matches  330;  Conservative   67;  Mismatches  131;  Indels   90;  Gaps   14;

Qy       587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
Db         3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy       647 RASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLTPSGAM 704
Db        59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTPSGEA 118

Qy       705 PNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 764
Db       119 PNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELREATSPKANKE 178

Qy       765 ILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWC 824
Db       179 ILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWC 238

Qy       825 MQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPI 884
Db       239 VQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAKQLGADEKEYHAEGGKVPI 298

Qy       885 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 944
Db       299 KWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPI 358

Qy       945 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFY 1003
Db       359 CTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFY 418

Qy      1004 RSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP 1063
Db       419 RTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------------- 449

Qy      1064 SEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPL 1118
Db       450 -----SRTPLLSSLSATSNNSATNCIDRNGG-----------H---------------- 476

Qy      1119 PSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERAKTLSPGKNG 1177
Db       477 PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAISKLPIDSRYQN- 527

Qy      1178 VVKDVFAFGGAVENPEYL 1195
Db       528 ------SHSTAVDNPEYL 539


RESULT 15
EGFR_CHICK
ID   EGFR_CHICK     STANDARD;      PRT;   703 AA.
AC   P13387;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE   (Fragment).
GN   EGFR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88261272; PubMed=3260329;
RA   Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA   Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT   "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT   expression in mouse cells, and differential binding of EGF and
RT   transforming growth factor alpha.";
RL   Mol. Cell. Biol. 8:1970-1978(1988).
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M20386; AAA48760.1; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 4.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     30
FT   CHAIN        31   >703       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    667       POTENTIAL.
FT   DOMAIN      668   >703       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    197    206       BY SIMILARITY.
FT   DISULFID    201    214       BY SIMILARITY.
FT   DISULFID    222    230       BY SIMILARITY.
FT   DISULFID    226    238       BY SIMILARITY.
FT   DISULFID    239    247       BY SIMILARITY.
FT   DISULFID    243    255       BY SIMILARITY.
FT   DISULFID    258    267       BY SIMILARITY.
FT   DISULFID    271    298       BY SIMILARITY.
FT   DISULFID    302    314       BY SIMILARITY.
FT   DISULFID    318    333       BY SIMILARITY.
FT   DISULFID    336    340       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    581       BY SIMILARITY.
FT   DISULFID    569    589       BY SIMILARITY.
FT   DISULFID    592    601       BY SIMILARITY.
FT   DISULFID    605    627       BY SIMILARITY.
FT   DISULFID    630    638       BY SIMILARITY.
FT   DISULFID    634    646       BY SIMILARITY.
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    200    200       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    359    359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    420    420       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    633    633       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    648    648       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     703    703
SQ   SEQUENCE   703 AA;  77427 MW;  AFF2DE11B735A690 CRC64;

  Query Match             22.3%;  Score 1518;  DB 1;  Length 703;
  Best Local Similarity   43.1%;  Pred. No. 9.1e-75;
  Matches  305;  Conservative  109;  Mismatches  263;  Indels   30;  Gaps   14;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |   :| ||  | |        :|| ||: ||      | |   |: :|  |:||  |||
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
                :||:  |  |:||: |||| |||||| | |  :||: |:|:|| |:::::||||| |
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
              :| |          ||||| :: |:||| ||| |  ||:||  ||:|| ||   : :  |

Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              |::|  :   :|  | | |    |||   ::||:||: :||  |: ||:| :|:|||| |
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||: ||||||  |      |:  || || ||  |::   ||||:|:|||:|| |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||   |     ||  |:| ::|:||   |::|| |:|:  :|      |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  | || | :: || :| ||  : | || |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
                ||  |:||::|||   :| ||| :  | |  |||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: |  ||     |      || :::|  :   |  ||:  ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
              |||::| :||| |||:     ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIV 655
              ||  ||:|| :   : :||: |    || |  |||  |     :|||     |   ||
Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 SAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLTP 700
              : || |:| :||:|:  |: :      |    ||        :|||||
Db        657 AGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLTP 703


Search completed: July 22, 2003, 09:18:21
Job time : 20.625 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.75 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-653-675-14
Perfect score: 6814
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6097	89.5	1259	6	O18735	O18735 canis famil
2	3060	44.9	1209	11	Q9QX70	Q9qx70 rattus norv
3	3038	44.6	1210	11	Q9EP98	Q9ep98 mus musculu
4	2647	38.8	1165	13	Q9YH40	Q9yh40 xiphophorus
5	2636.5	38.7	1137	13	Q9W6F6	Q9w6f6 gallus gall
6	2265	33.2	1328	13	P79754	P79754 fugu rubrip
7	2012.5	29.5	1433	5	Q9BIH9	Q9bih9 anopheles g
8	1871	27.5	419	4	Q9UK79	Q9uk79 homo sapien
9	1739	25.5	367	11	Q8R2X1	Q8r2x1 mus musculu
10	1697.5	24.9	412	4	Q8WYV0	Q8wyv0 homo sapien
11	1643	24.1	729	15	Q86712	Q86712 avian rous-
12	1641	24.1	567	15	Q86714	Q86714 avian rous-
13	1576.5	23.1	962	15	Q64895	Q64895 avian eryth
14	1568	23.0	545	15	Q85468	Q85468 avian eryth
15	1506.5	22.1	655	11	Q9WVF5	Q9wvf5 mus musculu
16	1490.5	21.9	643	11	Q9ERV6	Q9erv6 mus musculu
17	1268	18.6	1193	5	Q9Y1X8	Q9y1x8 ephydatia f
18	1180.5	17.3	1368	5	Q23821	Q23821 caenorhabdi
19	1142	16.8	1717	5	Q26566	Q26566 schistosoma
20	1126	16.5	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11	Q9ESE0	Q9ese0 rattus norv
22	942.5	13.8	599	13	Q9PSH2	Q9psh2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	887	13.0	176	11	Q923V5	Q923v5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8szw1 drosophila
27	754.5	11.1	311	13	Q99162	Q99162 xiphophorus
28	741.5	10.9	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
29	734	10.8	331	4	Q9BUD7	Q9bud7 homo sapien
30	725	10.6	1671	5	Q9NJV5	Q9njv5 biomphalari
31	723	10.6	149	6	Q9BG66	Q9bg66 oryctolagus
32	680.5	10.0	1368	13	Q8UW85	Q8uw85 paralichthy
33	680	10.0	1418	13	O93457	O93457 scophthalmu
34	673.5	9.9	1369	13	Q8UW86	Q8uw86 paralichthy
35	672	9.9	1412	13	Q8UW84	Q8uw84 paralichthy
36	660.5	9.7	1472	5	Q9U5A8	Q9u5a8 bombyx mori
37	658	9.7	1358	13	O73798	O73798 xenopus lae
38	640	9.4	1245	13	Q9YGH8	Q9ygh8 scophthalmu
39	636.5	9.3	1418	13	Q8UW83	Q8uw83 paralichthy
40	630.5	9.3	2144	5	Q9VD94	Q9vd94 drosophila
41	624	9.2	1371	11	Q9QVW4	Q9qvw4 rattus sp.
42	602.5	8.8	1091	4	Q9UMQ4	Q9umq4 homo sapien
43	595.5	8.7	1072	4	Q9BTB0	Q9btb0 homo sapien
44	595	8.7	987	11	Q91YM0	Q91ym0 mus musculu
45	586	8.6	1036	4	Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErbB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB008451; BAA23127.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

```
  Query Match             89.5%;  Score 6097;  DB 6;  Length 1259;
  Best Local Similarity   89.1%;  Pred. No. 0;
  Matches 1126;  Conservative   48;  Mismatches   74;  Indels   16;  Gaps    4;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGLLLALLPSGAAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| |||||||||||||||||||:|||:|||||||||||||||||||||||||||
Db         61 ELTYLPANASLSFLQDIQEVQGYVLIAHSQVRQIPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||    |  ||: |||||||||||||||||||:||||:||||||||:||||||||
Db        121 DPLEGGIPAPGAAQGGLRELQLRSLTEILKGGVLIQRSPQLCHQDTILWKDVFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||  || |||| || : ||| || |||||||||||||||||||| |||||||||
Db        181 LTLIDTNRFSACPPCSPACKDAHCWGASSGDCQSLTRTVCAGGCARCKGPQPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||:|||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||:||| ||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLRVFEALEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              :|||||||:|||||:||:|||||||||||||||||||||||||||| |   |||||||
Db        421 NLSVFQNLRVIRGRVLHDGAYSLTLQGLGISWLGLRSLRELGSGLALIHRNARLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||:||||:||||||||||:  || ||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHSANRPEEECVGEGLACYP-CAHGHCWGPGPTQCVNCSQFLRGQEC 539

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |||||||||||||||  |:||||| |||||||||||||| ||||||||||||||||||||
Db        540 VEECRVLQGLPREYVKDRYCLPCHSECQPQNGSVTCFGSEADQCVACAHYKDPPFCVARC 599

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              |||||||||:|||||| |||| |||||||||||| |||:|||||||||:|||::||||
Db        600 PSGVKPDLSFMPIWKFADEEGTCQPCPINCTHSCADLDEKGCPAEQRASPVTSIIAAVVG 659

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              ||| ||:|:| |||          :|:   |:  ::    :||||||||||||||||||
Db        660 ILLAVVVGLVLGILIKRRRQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 715

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 775

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              |||||||||||||||||||||||||||||||:|||||||||||||:|||||||||
Db        776 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSYL 835

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIPP 895

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 955

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        956 KCWMIDSECRPRFRELVAEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1015

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              |||||||||||||||||:| ||||| | ||||||||:|||:|||||||||| |:||||
Db       1016 LVDAEEYLVPQQGFFCPEPTPGAGGTAHRRHRSSSTRNGGGELTLGLEPSEEEPPKSPLA 1075

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              |||||||||||||||||||||||||: |||||||||||||||| |||| ||||||||||
Db       1076 PSEGAGSDVFDGDLGMGAAKGLQSLPSQDPSPLQRYSEDPTVPLPPETDGKVAPLTCSPQ 1135

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLER-----AKTLSPGKNGVVKDVFAFGGAV 1189
              |||||||:| ||||   ||||| :|||||||||      ||||||||||||||||| ||
Db       1136 PEYVNQPEVWPQPPLALEGPLPPSRPAGATLERPKTLSPKTLSPGKNGVVKDVFAFGSAV 1195

Qy       1190 ENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGL 1249
              |||||| |:| |||||||||||||||||||||||| |||:|||||:|||||||||||
Db       1196 ENPEYLAPRGRAAPQPHPPPAFSPAFDNLYYWDQDPSERGSPPSTFEGTPTAENPEYLGL 1255

Qy       1250 DVPV 1253
              ||||
Db       1256 DVPV 1259
```

RESULT 2
Q9QX70
ID   Q9QX70   PRELIMINARY;   PRT; 1209 AA.
AC   Q9QX70;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor.
GN   EGFR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RX   MEDLINE=90258888; PubMed=2342466;
RA   Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA   Earp H.S.;
RT   "A truncated, secreted form of the epidermal growth factor receptor is
RT   encoded by an alternatively spliced transcript in normal rat tissue.";
RL   Mol. Cell. Biol. 10:2973-2982(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Petch L.A.;
RL   Submitted (NOV-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Guttridge K., Dawson T.L., Earp H.S.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M37394; AAF14008.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1209 AA; 134891 MW; 96FEE7F6CC1B7773 CRC64;

```
  Query Match             44.9%;  Score 3060;  DB 11;  Length 1209;
  Best Local Similarity   49.3%;  Pred. No. 8.9e-223;
```

```
  Matches  630;  Conservative  164;  Mismatches  366;  Indels  118;  Gaps   28;

Qy          3 LAALCRWGLLLALLPPGA-ASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |||||  |        ||     :|| ||  :|    : | |    |: ::  |:|| ||||
Db         15 LAALCAAG--------GALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMFNNCEVVLGNLE 66

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  ||||: |||| |||||| | | ::||: |:|:||  |:|: |||||| |
Db         67 ITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAVLSN-- 124

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
                  | :  |||||  :|:|  ||| | |     || ||  :|| |:||  ::  |:
Db        125 --------YGTNKTGLRELPMRNLQEILIGAVRFSNNPILCNMETIQWRDIV-QDVFLSN 175

Qy        182 TLIDTNRS-RACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              :|  |     |  | | |     |||    |:|| ||: :||   |: ||:|  |:|||| |
Db        176 MSMDVQRHLTGCPKCDPSCPNGSCWGRGEENCQKLTKIICAQQCSRRCRGRSPSDCCHNQ 235

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:|| |
Db        236 CAAGCTGPRESDCLVCHRFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKC 295

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              | ||: || |||  |     ||  |||  :|:||  || :|| |:|:   | ::  |
Db        296 PRNYVVTDHGSCVRACGPDYYEV-EEDGVSKCKKCDGPCRKVCNGIGIGEFK-DTLSINA 353

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  |  | | |  || :| ||  : | || | :|:: :|::||||:| | |||::
Db        354 TNIKHFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPEN 413

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
                ||  |:||::|||   :| :|| : || |: ||||||:|:  |  :|  | :||: :
Db        414 WTDLHAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYAN 473

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: | :||  |:|      || | :|      |: ||:   |||| || ||:|    ||:
Db        474 TINWKKLFGTPNQKTKIMNNRAEKDCKATNHVCNPLCSSEGCWGPEPTDCVSCQNVSRGR 533

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||::| :|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||
Db        534 ECVDKCNILEGEPREFVENSECIQCHPECLPQTMNITCTGRGPDNCIKCAHYVDGPHCVK 593

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP-LTSIVSA 657
              ||||:  : : : :||| |    |  |  |||: |     |||  :|   | : || :
Db        594 TCPSGIMGENNTL-VWKFADANNVCHLCHANCTYGCAGPGLKGC--QQPEGPKIPSIATG 650

Qy        658 VVGILLVVVLGVVFGILIF---NNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              :|| || :|: |  || :|           ||        :|||||  |||| :||||
Db        651 IVGGLLFIVV-VALGIGLFMRRRQLVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILK 709

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||| :|:|||||||||||||:|||:|| ||||||||  ||| |||||||||||||||||
Db        710 ETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMAS 769

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              | :|:|  |||||||||||||:|||||||||||:|||::  :||| |||||:||||||:||
Db        770 VDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYL 829

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              || |||||||||||||||:| ||||||||||:||  :| ||||:|||||||||||||||
Db        830 EDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILH 889

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              | :|||||||||||||||||||:||||||||  ||  :|||||||||||||||||||||
Db        890 RIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMV 949

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMG 1013
              ||||||:: ||:||||: |||:|||||||::||| :| :    || || |||:|:|:::||
Db        950 KCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMEEEDME 1009

Qy       1014 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1073
              |:|||:|||:||||||                  | ||                |:||
Db       1010 DVVDADEYLIPQQGFF----------------NSPST----------------SRTPL 1035

Qy       1074 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTC 1131
                |  | |:       :       |    : : ||||| |||  |  :   | ::
Db       1036 LSSLSANSN----SSTVACINRNGSCRVKEDAFLQRYSSDPTSVLTEDNIDDTFL----- 1086

Qy       1132 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1191
                | |||:||  |  :|    : |:   :|              :||::  ::      || |
Db       1087 -PVPEYINQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVSN 1134

Qy       1192 PEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPST 1234
              |||| | |          |   |  ||:   | |         :|       |:   |
Db       1135 PEYLNTAQ---------PTCLSSGFDSSALWIQKGSHQMSLDNPDYQQDFFPKEAKPNGI 1185

Qy       1235 FKGTPTAENPEYLGLDVP 1252
              ||| ||||| ||| :  |
Db       1186 FKG-PTAENAEYLRVAPP 1202
```

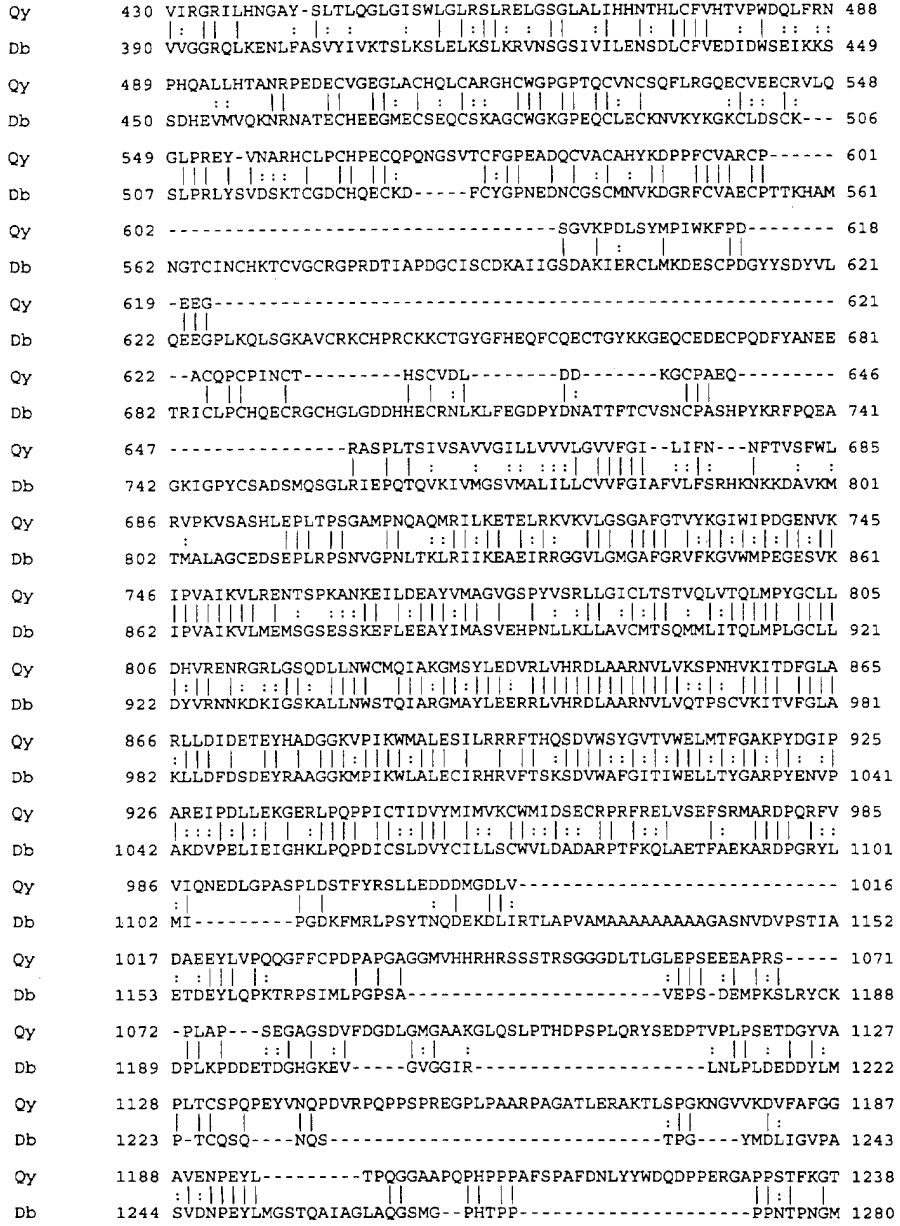
RESULT 3
Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maihle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maihle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.

```
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase.
SQ   SEQUENCE   1210 AA;  134840 MW;  62CD021C9DE32E18 CRC64;

  Query Match             44.6%;  Score 3038;  DB 11;  Length 1210;
  Best Local Similarity   48.9%;  Pred. No. 4.1e-221;
  Matches  622;  Conservative  167;  Mismatches  370;  Indels  114;  Gaps   25;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 LVVVLGVVFGILIF---NNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILKETELR 719
Db        657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFK 715

Qy        720 KVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPY 779
Db        716 KIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPH 775

Qy        780 VSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRL 839
Db        776 VCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRL 835

Qy        840 VHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTH 899
Db        836 VHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTH 895

Qy        900 QSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMI 959
Db        896 QSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVKCWMI 955

Qy        960 DSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDA 1018
Db        956 DADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVVDA 1015

Qy       1019 EEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEG 1078
Db       1016 DEYLTPQQGFF----------------NSPST-----------------SRTPLLSSLS 1041

Qy       1079 AGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPE 1136
Db       1042 ATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------PVPE 1091

Qy       1137 YVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL- 1195
Db       1092 YVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNPEYLN 1140

Qy       1196 TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKGTP 1239
Db       1141 TAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIFKG-P 1190

Qy       1240 TAENPEYLGLDVP 1252
Db       1191 TAENAEYLRVAPP 1203


RESULT 4
Q9YH40
ID   Q9YH40      PRELIMINARY;      PRT;  1165 AA.
AC   Q9YH40;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase proto-oncogene.
GN   XMRK.
OS   Xiphophorus xiphidium.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8086;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RX   MEDLINE=98241172; PubMed=9582016;
RA   Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA   Altschmied J., Schartl M.;
RT   "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT   overexpression and mutational alterations.";
RL   Oncogene 16:1681-1690(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53471; AAD10500.2; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
```

```
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1165 AA;  129614 MW;  7F7EE38D8771A74E CRC64;

  Query Match             38.8%;  Score 2647;  DB 13;  Length 1165;
  Best Local Similarity   44.6%;  Pred. No. 1.7e-191;
  Matches  571;  Conservative  164;  Mismatches  388;  Indels  156;  Gaps   32;

Qy          1 MELAALCRWGLLLALLPPG-AAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQV 55
              :||  |    ||| ||  |   ||    :|| ||  :: :    : |   :: :| || |
Db          4 LELLEL----LLLLLLSIGRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNV 56

Qy         56 VQGNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALA 115
              |  |||:||    |  ||||| |||| |||||| |:|  :|| ||::||  |:| |: |
Db         57 VLENLEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLL 116

Qy        116 VLDNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHK 175
              |: |    | ::|     ||::||| :||||| ||| :  || ||  :|| | ||  |
Db        117 VMSNYQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDK 173

Qy        176 NNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTD 234
              :   : ||     | |  | | |    ||      ||  |: :||  |  ||:|| | |
Db        174 TSNPTMNLIPHAFERQCQKCDPGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPID 233

Qy        235 CCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
              ||:| || ||||||: :|||||  ||  | |:  ||     |:  : : : ||  :|||||:
Db        234 CCNEHCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAA 293

Qy        295 CVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGM----QYI 350
              ||  || ||: |: |:|    |      ||  |:| : |: |   | :|| |:|:      |
Db        294 CVKECPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSLSNTI 351

Qy        351 KANSKFIGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
                ||  ||     |: | || | :     ||:|||     |: || |    |::||||||
Db        352 AVNSTNIG----SFSNCTKINGDIILNRNSFEGDPHYKIGPMDPEHLWNLTTVKEITGYL 407

Qy        411 YISAWPDSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIH 469
              |  ||::: |||||||::||||    :  :| : :|   : |||| ||:|: :|  ::
Db        408 VIMWWPENMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVSHLQWLGLSSLKEVSAGNVILK 467

Qy        470 HNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCV 529
              :   | :  |: | :|||:  |:: : |         |   |:  |:   |||||| ||
Db        468 NTPQLRYASTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCV 520

Qy        530 NCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAH 589
              :|     ||  ||  | :||| |||       |: || ||  |  |:||:||    |  |||
Db        521 SCLHVDRGGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKCAH 580

Qy        590 YKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS 649
              ::| | |: ||| |:  |   : |||: |: | ||||  ||| |        ||  :  |
Db        581 FQDGPQCIPRCPHGMLGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGD-IVS 638

Qy        650 PLTSIVSAVVGILLVVVLGVVFGIL-----IFNNFTVSFWLRVPKVSASHLEPLTPSGAM 704
              :  | | |:|: |:: ::  :|    |   |:   |:  ::    :|||||||
Db        639 HSSLAVGLVSGLLITVIVALLIVVLLRRRRIKRKRTIRRLLQEKEL----VEPLTPSGQA 694

Qy        705 PNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 764
              |||| :|||||| :| :|||||||||||||:| |||||::|||||||||| |||| |:|
Db        695 PNQAFLRILKETEFKKDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQE 754

Qy        765 ILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWC 824
              :|||||||  |   |:| |||||||||  |||||||||||||||:||::: |:  | |||||
Db        755 VLDEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWC 814

Qy        825 MQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPI 884
              :||||||:|||:  |||||||||||:|:|||||||||||::||  || || | ||||||
Db        815 VQIAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQAHGGKVPI 874

Qy        885 KWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPI 944
              |||||||||:  :|||||||||||||||||||:|||||||||:||  :|| ||||||||
Db        875 KWMALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPI 934

Qy        945 CTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYR 1004
              |||:||||::|||||   |||||||| |||:|||| |::|||       || |   :
Db        935 CTIEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSPSDRRLFS 991

Qy       1005 SLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPS 1064
              ||  ||  |:|||:|||:                     |::  : | |
Db        992 RLLSSDD--DVVDADEYLL--------------------RYKRIN-RQGS---------- 1018

Qy       1065 EEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETD 1123
                    | | |                          | : |   || |||   | : |
Db       1019 ------EPCIPPNGH------------------PVRENSIALRYISDPTQNALEKDLD 1052

Qy       1124 GYVAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKT 1170
              |:          ||||||             |:  :     ||:  ::: |      | :
Db       1053 GH----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEY 1102

Qy       1171 LSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGA 1230
              |:  :| :         |::::|:|   | |          :  ||           |: ||
Db       1103 LNTNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGA 1137

Qy       1231 PPSTFKGTPTAENPEYLGL 1249
                    | ||| |||||
Db       1138 LTGNGMFLPAAENLEYLGL 1156


RESULT 5
Q9W6F6
ID   Q9W6F6      PRELIMINARY;      PRT;  1137 AA.
AC   Q9W6F6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase (Fragment).
GN   ERBB4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HINDBRAIN;
RX   MEDLINE=99263203; PubMed=10328884;
RA   Dixon M., Lumsden A.;
RT   "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
RT   embryonic chick hindbrain.";
RL   Mol. Cell. Neurosci. 13:237-258(1999).
DR   EMBL; AF121963; AAD31764.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001368; TNFR_c6.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 1.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW   Kinase; Tyrosine-protein kinase.
FT   NON_TER       1       1
SQ   SEQUENCE   1137 AA;  127927 MW;  4D616436F87DC84F CRC64;

  Query Match             38.7%;  Score 2636.5;  DB 13;  Length 1137;
```

```
  Best Local Similarity   45.9%;  Pred. No. 1e-190;
  Matches  525;  Conservative  172;  Mismatches  357;  Indels   91;  Gaps   26;

Qy        161 LCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVC 220
              ||: ||| |:||          ||: || | |  |  |  | | |||| :   ||:||:|||
Db          3 LCFADTIHWQDIVRNPWASNFTLVPTNGSSGCGRCHKSCTG-RCWGPTENHCQTLTKTVC 61

Qy        221 AGGC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTF 279
              |  |  || ||  :|||| :|| ||:||| :|| ||::|| || |   ||    ||  ||
Db         62 AEQCDGRCYGPYVSDCCHRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTF 121

Qy        280 ESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCA 339
              :   |   :||:|| ||  ||:|:: |  ||  ,||     ||  |:| : |: |:  |
Db        122 QLEHNHNAKYTYGAFCVKKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICP 179

Qy        340 RVCYGLGMQYIKANSKFIGITELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQ 398
              : | |:|   :  ::: :  : :: |  | || |:| ||      |||     | : ||:|
Db        180 KACDGIGTGSL-VSAQTVDSSNIDKFINCTKINGNLIFLVTGIHGDPYHTIAAINPEKLN 238

Qy        399 VFETLEEITGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSL 458
              :|:|: |||||| | :||::: |  || ||  | || |::|   | |:  ||: |  :||
Db        239 IFQTVREITGYLNIQSWPENMTDFRVFSNLVTIGGRALYSGLSLLILKQQGITSLQFQSL 298

Qy        459 RELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARG 518
              ::: :|   |  |::||: ||| |  ||  | |  :   |:  :  |  :|: |::||:
Db        299 KQISAGNIYITDNSNLCYYHTVNWTSLFSTPSQKTVIHRNKKAENCTADGMVCNELCSSD 358

Qy        519 HCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCF 577
              |||||| ||::| :|:||: |:| | :  |  ||: |   |: | |:|:  ::  :||:
Db        359 GCWGPGPDQCLSCKRFIRGRTCIESCNLYDGEFREFANGSVCMECDPQCEKMEDNMITCY 418

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |  | |:|| | || :|| |::   |:  |:|: ||:  | ||  |||  |
Db        419 GPGPDHCTKCFHFKDGPNCVEKCPDGLQGANSF--IFKYADEDRECHPCHPNCTQGCRGP 476

Qy        638 DDKGC-----------PAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTVSFWL 685
                  |           |   | :||  | : |: |: ::|::|: | : :
Db        477 ASHDCIYYPWTRQSTLPQHAR-TPL--IAAGVIGGLFIIVIMGLTFAVYVRRK------- 526

Qy        686 RVPKVSASH-------LEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWI 738
              : |  |          :|||||||  |||||:||||||||::|||||||||||||||||:
Db        527 SIKKKRALRRFLETELVEPLTPSGTAPNQAQLRILKETELKRVKVLGSGAFGTVYKGIWV 586

Qy        739 PDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQL 798
              |:|| ||||||||:| | | |||| | :||| :|| :  |:: ||||:||: |:||||||
Db        587 PEGETVKIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQL 646

Qy        799 MPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVK 858
              ||:|||||:| |::  :||| |||||:|||||| ||: ||||||||||||||||||||||
Db        647 MPHGCLLDYVHEHKDNIGSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVK 706

Qy        859 ITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGA 918
              |||||||||: || ||:|||||:|||||||| |  |:||||||||||||:|||||||
Db        707 ITDFGLARLLEGDEKEYNADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGG 766

Qy        919 KPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMA 978
              ||||||| |||||||||||||||||||||||||:|||||||:: ||:|:|| :||||||
Db        767 KPYDGIPTREIPDLLEKGERLPQPPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMA 826

Qy        979 RDPQRFVVIQNED-LGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGA 1037
              |||||::||| :| :   || || |:::||:::|: |::|||||||| | |  | |   :
Db        827 RDPQRYLVIQGDDRMKLPSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTS 885

Qy       1038 GGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS--PLAP-SEGAGSDVFDGDLGMGAAK 1094
                :     | ||     |: |    ||    | || ::|| :::|:     |  :
Db        886 RTRIDSNRNQFVYRDGGYAAEQGV-PMPYRAPGCIIPEAPVAQGATAEIFEDTCCNGTLR 944

Qy       1095 GLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQP 1147
                :    : |  |||| |||| :|         : |||: |:   |: :|:|  :  |
Db        945 KQVATLAKEDSSTQRYSADPTVFIPERVIRGELDEDGYMTPMRDKPKTDYLNPVEENPFV 1004

Qy       1148 PSPREGPLPAA-RPAGATLERAKTLSPGKNGVVKDVF--------AFGGAVENPEYLTPQ 1198
                : | | |   |           :|      :| :         |   :|| |||

Db       1005 SRRKNGDLQAVDNPEYHN-------APNGQPKAEDEYVNEPLYLNTFANTLENAEYL--- 1054

Qy       1199 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGA--PPSTFKGTPT-------------AE 1242
                    : : |      ||||  ||:   | |      |   :  |              ||
Db       1055 -----KNNLPEKAKKAFDNPDYWNHSLPPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAE 1109

Qy       1243 NPEYL 1247
              |||||
Db       1110 NPEYL 1114


RESULT 6
P79754
ID   P79754      PRELIMINARY;      PRT;  1328 AA.
AC   P79754;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ErbB3.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99177347; PubMed=10077531;
RA   Gellner K., Brenner S.;
RT   "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT   rubripes.";
RL   Genome Res. 9:251-258(1999).
DR   EMBL; AF056116; AAC34391.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   1328 AA;  148613 MW;  A33303925 8B647E9 CRC64;

  Query Match             33.2%;  Score 2265;  DB 13;  Length 1328;
  Best Local Similarity   39.8%;  Pred. No. 1.8e-162;
  Matches  513;  Conservative  160;  Mismatches  418;  Indels  198;  Gaps   33;

Qy          9 WGLLLALLPP--GAASTQ----VCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLEL 62
              | |:|  :     |||:|    || ||  |   | |   :: :  |:||::: ||||:
Db          4 WRLILMCVASRLRAASSQTQEAVCPGTQNGLSSTGSQENQYNLNKDRYKGCEIIMGNLEI 63

Qy         63 TYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDP 122
              | : :|   |||: |:|| |||||| |  :::|| :||::||  |:|  :||:|  |
Db         64 TQIESNWDFSFLKTIREVTGYVLIAMNHFQEIPLGQLRVIRGNSLYERRFALSVFLN--- 120

Qy        123 LNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALT 182
                  |  |  | || :| | :|||| ||| |  |  | |   : |:||  :||   :
Db        121 ----YPKDG--PSGLNQLGLMNLTEILDGGVQIINNKYLRYGPWVYWRDII-RNNDAPIE 173

Qy        183 LIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCA 241
              :           ||      | |: |||   : || ||:||||  |  || |  | |||| :||
Db        174 IQFNGERGVCH---KSC-GNYCWGPGKDQCQILTKTVCAPQCNDRCFGTSPRDCCHIECA 229

Qy        242 AGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY 301
              ||| ||  :|| ||  || || |   ||  : ||  ||:   ||  :| :|: ||: ||
Db        230 AGCKGPLDTDCFACRLFNDSGACVPQCPQTLIYNKQTFQMETNPNAKYQYGSICVSQCPT 289
```

```
Qy        302 NYLSTDVGSCTLVCPLHNQEVTAEDGTQR-CEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              :::  |  ||  |||    ||  | |:|| || ||  | :|| | |  :      : :  :
Db        290 HFV-VDGSSCVSVCPPDKMEV--ERGSQRQCELCSGLCPKVCEGTGAE----QRQTVDSS 342

Qy        361 ELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
               :: |  | || ||| ||     ||   |  ||  ::|:|| |: |||  | | :||  |
Db        343 NIDSFINCTKIQGSLHFLVTGILGDDFKNVPPLDAKKLEVFRTVREITDILNIQSWPKEL 402

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              |||||  :|  |:|| |      : ::   :: ||||||||:  |   |  | |||: ||
Db        403 NDLSVFSSLTTIQGRSLFKRFSLMVMRIPTLTSLGLRSLREISDGSVYISQNAHLCYHHT 462

Qy        480 VPWDQLFRNPH-QALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              | | ||||    :|    :|||  ||| :|  |  ||:   |||||| ||::|  : |
Db        463 VNWTQLFRGSRVRANSLNSNRPMAECVADGRVCDPLCSDSGCWGPGPDQCLSCRNYSRHG 522

Qy        539 ECVEECRVLQGLPREYVNARH-CLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCV 597
              ||  |    |:|||:      |: |||||:|| |  :| || ||:|:||  ::| |:|:
Db        523 TCVAGCHFNSGIPREFAGLNGVCVACHPECKPQTGKASCTGPGADECMACTKFRDGPYCM 582

Qy        598 ARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSA 657
              : ||:||  |     |:|||: || |:||  ||| |       |     |: ||
Db        583 SSCPAGVN-DGEKGLIFKFPNREGHCEPCHQNCTQGCSGPGLNDC---LEAARLTISSGQ 638

Qy        658 VVGILLVVVLGVVFGILIFNNFTVSFWLRVPKV----------SASHLEPLTPSGAMPN 706
              : || | |  |::| :::|  |     : |   :          |   ||| | |
Db        639 ITGIALGVPAGLIFCLVLF--FLGMLYHRGLAIRRKRAMRRYLESGESFEPLGP-GEKGT 695

Qy        707 QAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIL 766
              :   |||| ::|||:| |||| |||| || |||:|| |||||||| :::::  :   ||
Db        696 KVHARILKPSDLRKIKPLGSGVFGTVSKGFWIPEGETVKIPVAIKTIQDSSGRQTFTEIT 755

Qy        767 DEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQ 826
              |    |  :  ||: |||||  : :||||||  :| ||:|:|:::  |  | |||||:|
Db        756 DHLLSMGSLDHPYIVRLLGICPGTCLQLVTQLSSHGSLLEHIRQHKTSLDPQRLLNWCVQ 815

Qy        827 IAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKW 886
              ||||| |||: |:||:|||||:|:|:   |:|:|:|:| ||  |: :|      | ||||
Db        816 IAKGMYYLEEHRVVHKNLAARNILLKNDYQVQISDYGVADLLYPDDKKYVYSETKTPIKW 875

Qy        887 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 946
              ||||||| ||:|||||||||||||:|:||:||  :  :|:| :|||||| || |||
Db        876 MALESILFRRYTHQSDVWSYGVTVWEMMSFGAEPYASVQPQEVPSVLEKGERLSQPAICT 935

Qy        947 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1006
              |||||:|||||||   || |:|| |:|:|||||| |::||: |
Db        936 IDVYMVMVKCWMIDENIRPTFKELASDFTRMARDPPRYLVIRMEG--------------- 980

Qy       1007 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1066
               ||  ||: :                          |  | |   | |   ||  ||
Db        981 -EDSGMGEFL----------------------------RRGSER---GLLEADLEEDEE 1007

Qy       1067 EAPRSPLAPSEGAGSDVFDGDLGMG---AAKGLQSLPTHDPSPLQ----------RYSED 1113
              |                          |:|   |   ||  |:   || |           ||
Db       1008 E--------------------GLGDRFATPSLQPSPSWSTSPSQINSYMVMTQLRYD-- 1044

Qy       1114 PTVPLPSETDGYVAPLTCSPQP-EYVNQ--------------PDVRPQPPSPREGPL--P 1156
                    |::  |  || | : : |                 ||      | ||  |
Db       1045 ----FAVSQGGHIGYLPMSPSPVDTIRQLWYQRSRLSSVRTLPDRSAFRRSSREAELCED 1100

Qy       1157 AARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFD 1216
               |: ||    |                 ||    |   |||
Db       1101 GAQCAGIFRVR-----------------FGSERGN-----PQGG--------------- 1122

Qy       1217 NLYYWDQDPPERGAPPSTFKGTPTAENPE 1245
                    |      : ||:||     |  |
Db       1123 -----QQRKLSTASSPSSFKTWAADEEDE 1146


RESULT 7
Q9BIH9
ID   Q9BIH9      PRELIMINARY;      PRT;  1433 AA.
AC   Q9BIH9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative epidermal growth factor receptor (Fragment).
GN   EGFR.
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC   Anopheles.
OX   NCBI_TaxID=7165;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SUA;
RA   Lycett G.J.;
RT   "Cloning, expression and localisation of the Anopheles gambiae
RT   epidermal growth factor receptor.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ301655; CAC35008.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 7.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Receptor.
FT   NON_TER       1       1
SQ   SEQUENCE   1433 AA;  159585 MW;  E3D9D88967724F07 CRC64;

  Query Match           29.5%;  Score 2012.5;  DB 5;  Length 1433;
  Best Local Similarity  32.3%;  Pred. No. 2.8e-143;
  Matches  468;  Conservative  202;  Mismatches  390;  Indels  387;  Gaps   38;

Qy         26 CTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTNASLSFLQDIQEVQGYVL 85
              | ||: :: :||: | |   ||  |  |  | ||||:|::      |:||| |:|| ||||
Db          1 CIGTNGRMSVPANREYHYKNLRDRYTNCTYVDGNLEITWIQNITDLNFLQHIREVTGYVL 60

Qy         86 IAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140
              |:   :  || | ||:|:||    |      |: | | |                |   : |
Db         61 ISLYDLPQVILPRLQIIRGRTTFKLNKWEEAYGLFV----------------SFSHMNTL 104

Qy        141 QLRSLTEILKGGVLIQRNPQLCYQDTILWKDI-FHKNNQLALTLIDTNRSRACHPCSPMC 199
              :| :| :|| | |    |  ||:  :| |::|       :  |   ::  | | || | |
Db        105 ELPALRDILGGSVGFFNNYNLCHMKSINWEEILLAPQTSMQYTFNFSSPERVCPPCHPSC 164

Qy        200 KGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDCLACLH 257
              :   |||| : :||  ::  |:  |:  || || | :|||  || |||||  |||||| :
Db        165 EVG-CWGEGAHNCQRFSKLNCSPQCSQGRCFGPKPRECCHLFCAGGCTGPTQSDCLACKN 223

Qy        258 FNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPL 317
              |   |:|:  || :  ||   :   |||:|:| :||:||  ||  :|  | ||:|   ||
Db        224 FYDDGVCKQECPPMQIYNPTNYFWEPNPDGKYAYGATCVRKCP-EHLLKDNGACVRKCPK 282

Qy        318 HNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELEFAGCKKIFGSLAFL 377
                    :|    | |   | : | | |:     :|  ||      :  |  | |||  |
Db        283 GKMPQNSE-----CVPCKGVCPKTCPGEGI----VHSDNIG----NYKDCTIIEGSLEIL 329

Qy        378 PESFDGDPASNT--------APLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQNLQ 429
              :||||      |          : |::|:|| |::||||:: | |   :   |: |:||:
Db        330 DQSFDGFQQVYTNFSFGPRYIKIDPDRLEVFSTVKEITGFINIQAHHPNFTTLNYFRNLE 389
```



RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 2.4e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027080; AAH27080.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   367 AA;  40163 MW;  0BE03395F9E101B0 CRC64;

  Query Match             25.5%;  Score 1739;  DB 11;  Length 367;
  Best Local Similarity   88.0%;  Pred. No. 2e-123;
  Matches  323;  Conservative   11;  Mismatches   33;  Indels    0;  Gaps    0;

Qy        887 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        947 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||:||:||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSL 120

Qy       1007 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1066
              |||||||:|||||||||||||||| |||| | |    | |||||| |||||:||||||||||
Db        121 LEDDDMGELVDAEEYLVPQQGFFSPDPALGTGSTAHRRHRSSSARSGGGELTLGLEPSEE 180

Qy       1067 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1126
              | |||||||||||||||||||| :|  ||||||  || ||||||||||:||| ||||||
Db        181 EPPRSPLAPSEGAGSDVFDGDLAVGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYV 240

Qy       1127 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1186
              ||| ||||||||||:|||| |   ||| |  |||||||| ||||||||||||||||||||
Db        241 APLACSPQPEYVNQPEVRPQSPLTPEGPPPPIRPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1187 GAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1246
              |||||||| |: | | |||| ||||||||||||||:  |:| |||||:||||||||||
Db        301 GAVENPEYLAPRAGTASQPHPSPAFSPAFDNLYYWDQNSSEQGPPPSTFEGTPTAENPEY 360

Qy       1247 LGLDVPV 1253
              |||||||
Db        361 LGLDVPV 367


RESULT 10
Q8WYV0
ID   Q8WYV0      PRELIMINARY;      PRT;   412 AA.
AC   Q8WYV0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 44.7 kDa protein.
GN   PP3659.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA   Wan D.F., Gu J.R.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF318349; AAL55856.1; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   412 AA;  44702 MW;  034397FF3F27D2BC CRC64;

  Query Match             24.9%;  Score 1697.5;  DB 4;  Length 412;
  Best Local Similarity   80.5%;  Pred. No. 3.3e-120;
  Matches  330;  Conservative    5;  Mismatches   22;  Indels   53;  Gaps    4;

Qy        887 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        947 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 120

Qy       1007 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 180

Qy       1067 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 240

Qy       1127 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1186
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        241 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1187 GAVENPEYLTPQGGAAPQP------------------------HPPPA---FSPAFDNL 1218
              |||||||||||||||| |                        || ||:       :
Db        301 GAVENPEYLTPQGGAALSPTLLLPSAQPSTTSITGTRTHQSGGLHPAPSKGHLRQRTQST 360

Qy       1219 YYWD-QDPPER-----------------------GAPPSTFKGTPTAEN 1243
              : |  |  ||                       ||:| :||  | |
Db        361 WVWTCQCEPEGQVRRSPDVSSGSREGLTSAGIKRWEGPPTTSRGTCHARN 410


RESULT 11
Q86712
ID   Q86712      PRELIMINARY;      PRT;   729 AA.
AC   Q86712;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyprotein.
GN   POLYPROTEIN.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60725.1; -.
DR   HSSP; P03322; 1A6S.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR004028; Retro_M.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02813; Retro_M; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   729 AA;  80649 MW;  84D2F6914EFE1D63 CRC64;

  Query Match             24.1%;  Score 1643;  DB 15;  Length 729;
  Best Local Similarity   53.8%;  Pred. No. 1e-115;
```

```
  Matches  344;  Conservative   71;  Mismatches  124;  Indels  100;  Gaps   14;

Qy        582 DQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKG 641
              | |: |||: | | ||  ||:||  :   : :||: |    || |  ||| |     :|
Db        152 DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEG 210

Qy        642 CPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLT 699
              ||     |   || : || |:| :||:|:  |: :     |    ||       :||||
Db        211 CP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLT 267

Qy        700 PSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSP 759
              |||  |||| :|||||| :|||||||||||||||:|||:|| |||||||| ||| |||
Db        268 PSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSP 327

Qy        760 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 819
              |||||||||||||| | :| | ||||||||||||||:||||||||||::||::  :|||
Db        328 KANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQY 387

Qy        820 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 879
              |||||:||||||:|||: |||||||||||||||:| ||||||||||:||  || ||||:|
Db        388 LLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEG 447

Qy        880 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 939
              |||||||||||||| | :|||||||||||||||||||:||||||| ||  :|||||||
Db        448 GKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERL 507

Qy        940 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPL 998
              |||||||||||||||||||:: ||:||||::|||:|||||  |::||| :| :   ||
Db        508 PQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPT 567

Qy        999 DSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLT 1058
              || |||:|:|::|| |:|||:|||||  ||||                | ||
Db        568 DSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-------- 603

Qy       1059 LGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSED 1113
                        |:||     | |  : ::  |        :  |  |  : | :|||| |
Db        604 ----------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSD 645

Qy       1114 PTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTL 1171
              ||      |:  ||::      | ||||||  : |: ||
Db        646 PTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS----------------------T 676

Qy       1172 SPGKNGVVKDVF---------------AFGGAVENPEYL 1195
              :  :| :  ::                :   ||:|||||
Db        677 AMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 715


RESULT 12
Q86714
ID   Q86714      PRELIMINARY;      PRT;   567 AA.
AC   Q86714;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   V-erbB protein (Fragment).
GN   V-ERBB.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60727.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Tyrosine-protein kinase.
FT   NON_TER       1       1
SQ   SEQUENCE   567 AA;  63390 MW;  C6D9CBA7ADF725E1 CRC64;

  Query Match             24.1%;  Score 1641;  DB 15;  Length 567;
  Best Local Similarity   53.8%;  Pred. No. 1e-115;
  Matches  344;  Conservative   71;  Mismatches  124;  Indels  100;  Gaps   14;

Qy        582 DQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKG 641
              | |: |||: | | ||  ||:||  :   : :||: |    || |  ||| |     :|
Db          3 DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEG 61

Qy        642 CPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHLEPLT 699
              ||     |   || : || |:| :||:|:  |: :     |    ||       :||||
Db         62 CP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELVEPLT 118

Qy        700 PSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSP 759
              |||  |||| :|||||| :|||||||||||||||:|||:|| |||||||| ||| |||
Db        119 PSGEAPNQAHLRILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSP 178

Qy        760 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 819
              |||||||||||||| | :| | ||||||||||||||:||||||||||::||::  :|||
Db        179 KANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQY 238

Qy        820 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 879
              |||||:||||||:|||: |||||||||||||||:| ||||||||||:||  || ||||:|
Db        239 LLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEG 298

Qy        880 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 939
              |||||||||||||| | :|||||||||||||||||||:||||||| ||  :|||||||
Db        299 GKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERL 358

Qy        940 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPL 998
              |||||||||||||||||||:: ||:||||::|||:|||||  |::||| :| :   ||
Db        359 PQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPT 418

Qy        999 DSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLT 1058
              || |||:|:|::|| |:|||:|||||  ||||                | ||
Db        419 DSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-------- 454

Qy       1059 LGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSED 1113
                        |:||     | |  : ::  |        :  |  |  : | :|||| |
Db        455 ----------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSD 496

Qy       1114 PTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTL 1171
              ||      |:  ||::      | ||||||  : |: ||
Db        497 PTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS----------------------T 527

Qy       1172 SPGKNGVVKDVF---------------AFGGAVENPEYL 1195
              :  :| :  ::                :   ||:|||||
Db        528 AMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 566


RESULT 13
Q64895
ID   Q64895      PRELIMINARY;      PRT;   962 AA.
AC   Q64895;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Gag,v-erb-A,v-erb-B protein.
GN   GAG,V-ERB-A,V-ERB-B.
OS   Avian erythroblastosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=90206603; PubMed=1969616;
RA   Bruskin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT   "Six amino acids from the retroviral gene gag greatly enhance the
RT   transforming potential of the oncogene v-erb-B.";
RL   Oncogene 5:15-24(1990).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; X52209; CAA36459.1; -.
DR   EMBL; X52211; CAA36459.1; JOINED.
DR   HSSP; P10828; 2NLL.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW   Transcription regulation; Transferase; Tyrosine-protein kinase;
KW   Zinc-finger.
SQ   SEQUENCE   962 AA;  108320 MW;  3C5AED791E4E95CE CRC64;

  Query Match             23.1%;  Score 1576.5;  DB 15;  Length 962;
  Best Local Similarity   50.2%;  Pred. No. 1.7e-110;
  Matches  347;  Conservative   71;  Mismatches  154;  Indels  119;  Gaps   18;

Qy        541 VEECRVLQGLPRE-YVNAR-HCLP------------------------------CHPECQ 568
              :|:|:    |  | |:| | | :|                              |  |
Db        354 IEKCQESYLLAFEHYINYRKHNIPHFWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413

Qy        569 PQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPI 628
              ||       ||  | |: |||: | | ||  ||:||  :   : :||: |    || |
Db        414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHP 465

Qy        629 NCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLR 686
              |||  |     :|||      |   || : || |:| :||:|:  |: :     |    ||
Db        466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLR 522

Qy        687 VPKVSASHLEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKI 746
                :|||||||  |||| :||||||| :|||||| |||||||||:|||:|| | |
Db        523 RLLQERELVEPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTI 582

Qy        747 PVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLD 806
              ||||| ||| |||||||||||||||| | :|:| |||||||||||||:||||||||||
Db        583 PVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLD 642

Qy        807 HVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLAR 866
              ::||::  :||| ||||:||||||:|||:  :||||||||||||:| ||||||||||:
Db        643 YIREHKDNIGSQYLLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAK 702

Qy        867 LLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPA 926
              |  || ||||:|||||||||||||| | :|||||||||||||||||||:||||||||
Db        703 QLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPA 762

Qy        927 REIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVV 986
              ||  :|||||||||||||||||||||||  :: ||:||||::|||:|||||  |::|
Db        763 SEISSVLEKGERLPQPPICTIDVYMIMVKCWMSGADSRPKFRELIAEFSKMARDPPRYLV 822

Qy        987 IQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRH 1045
              || :| :  || || |||:|:|::|| |:|||:|||| ||||
Db        823 IQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF---------------- 866
```

```
Qy       1046 RSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPS 1105
                | ||                |:|| |  | |:       | | :     |
Db        867 NSPST-----------------SRTPLLSSLSATSN-------NSATKCIDRNGGH--- 898

Qy       1106 PLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT- 1164
                          |   ||::     | ||||||  : |: ||              |
Db        899 -------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTA 937

Qy       1165 LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1195
              : :    |  :|       :   ||:|||||
Db        938 ISKLPMDSRYQN-------SHSTAVDNPEYL 961


RESULT 14
Q85468
ID   Q85468      PRELIMINARY;      PRT;   545 AA.
AC   Q85468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS   Avian erythroblastosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88217326; PubMed=2897102;
RA   Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT   "Common site of mutation in the erbB gene of avian erythroblastosis
RT   virus mutants that are temperature sensitive for transformation.";
RL   Oncogene Res. 1:265-278(1987).
DR   EMBL; X06943; CAA30024.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   545 AA;  60899 MW;  140DCE8CCA0F8AF4 CRC64;

  Query Match             23.0%;  Score 1568;  DB 15;  Length 545;
  Best Local Similarity   53.3%;  Pred. No. 3.2e-110;
  Matches  334;  Conservative   68;  Mismatches  133;  Indels   92;  Gaps   15;

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |: |||: | | || ||:||  :   : :||: |    || |  |||  |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIFNNFTV-SFWLRVPKVSASHL 695
              :|||      |   || : || |:| :||:|:  |: :     |    ||        :
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRRHIVRKRTLRRLLQERELV 114

Qy        696 EPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRE 755
              |||||||  |||| :||||||| :|||||| |||||||||:|||:|| | |||||| |||
Db        115 EPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELRE 174

Qy        756 NTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRL 815
              |||||||||||||||| | :|:| ||||||||||||||:||||||||::||::  :
Db        175 ATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNI 234

Qy        816 GSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEY 875
              ||| |||||:||||||:|||:  |||||||||||||:|  |||||||||: |  || ||
Db        235 GSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQLGADEKEY 294

Qy        876 HADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 935
              ||:|||||||||||| | :|||||||||||||||||:|||||||| ||  :|||
Db        295 HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354
```

```
Qy        936 GERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGP 994
              ||||||||||||||||||||||| |:: ||:||||::|||:||||| |::||| :| :
Db        355 GERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHL 414

Qy        995 ASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGG 1054
              || || ||| |:|::|| |:|||:||||| |||| | ||
Db        415 PSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST---- 454

Qy       1055 GDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQR 1109
                          |:||     | |  : ::  | : |          |
Db        455 --------------SRTPLLSSLSATSNNSATNCIDRNGG------------H------- 481

Qy       1110 YSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERA 1168
                  |   ||::     | ||||||  : |: ||     :         | : :
Db        482 ---------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQIQIYNYISLTAISKL 524

Qy       1169 KTLSPGKNGVVKDVFAFGGAVENPEYL 1195
                |  :|      :   ||:|||||
Db        525 PMDSRYQN-------SHSTAVDNPEYL 544
```

```
RESULT 15
Q9WVF5
ID   Q9WVF5      PRELIMINARY;      PRT;   655 AA.
AC   Q9WVF5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor (Epidermal growth factor receptor
DE   isoform 3).
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT   Carboxy-Terminal Truncated Receptors.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative Egfr transcripts encoding truncated receptor
RT   isoforms.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AF124513; AAD44149.1; -.
DR   EMBL; AF275366; AAG28047.1; -.
DR   EMBL; AF275364; AAG28047.1; JOINED.
DR   EMBL; AF275365; AAG28047.1; JOINED.
DR   EMBL; AK004944; BAB23688.1; -.
DR   EMBL; AK004883; BAB23641.1; -.
DR   EMBL; AK004911; BAB23662.1; -.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR002174; Furin-like.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 3.
KW   Receptor.
SQ   SEQUENCE   655 AA;  72906 MW;  6B34063B1BC928CB CRC64;

  Query Match             22.1%;  Score 1506.5;  DB 11;  Length 655;
  Best Local Similarity   44.5%;  Pred. No. 1.9e-105;
  Matches  285;  Conservative   97;  Mismatches  233;  Indels   25;  Gaps    9;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
              || ||   | |   :|| ||  :|   : | |   |: :|  |:|| ||||:||:  |
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
                ||||: |||| |||||| | | ::||: |:|:||  |:|: ||||:| |
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
                | :  ||||| :|:| ||| | |    || ||  ||| |:||    |  |  : |
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
                :  :|  | | |     |||   |:|| ||: :||  |: ||:|  |:|||| |||||
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              ||||: |||| |  |     |:  || |: ||  |::   ||||:|:|||:||  || ||
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              : || |||   |     ||  ||| ::|:|| || :|| |:|:   | ::  |  | ::
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
               |  |  | |  || :| ||  :  | || | :|:: :|::||||:| | ||||:  ||
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
              |:||::|||   :| :|| : || |: ||||||:|:  |  :|  | :||: :|: |
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
              :||  |:|      || | :|      |: ||:   |||| |  ||:|    ||:||||
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              :| :|:| |||:|    |: |||| ||  ::|| |   | |: |||| | | ||  ||:
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC 642
              |:  : : : :||: |    |  |  |||: |     :||
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGC 636
```

```
Search completed: July 22, 2003, 09:24:29
Job time : 51.75 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.875 Seconds
(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-653-675-14
Perfect score: 6814
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6610	97.0	1255	21	AAY92620	Human heregulin 2
2	6610	97.0	1255	22	AAE12130	Human tyrosine kin
3	6610	97.0	1255	22	AAB60167	HER2 transgene pla
4	6610	97.0	1255	23	AAU74545	Human HER2 (ErbB2)
5	6604	96.9	1255	17	AAW01111	HER-2/neu protein.
6	6604	96.9	1255	20	AAW92406	Human HER-2/neu on
7	6604	96.9	1255	21	AAB21198	Human HER-2/neu pr
8	6604	96.9	1255	21	AAY84780	Amino acid sequenc
9	6604	96.9	1255	22	AAB85458	Human HER-2/neu pr
10	6604	96.9	1255	22	AAG88267	HER2/neu amino aci
11	6604	96.9	1255	23	AAE24067	Human Her-2 protei
12	6604	96.9	1255	23	AAE20479	Human Her-2/neu pr
13	6604	96.9	1255	23	AAM51143	Human Her-2/neu on
14	6604	96.9	1255	23	AAU77114	Human Her-2/neu po
15	6561	96.3	1433	14	AAR39568	Sequence of c-erbB
16	6440	94.5	1223	23	AAU98923	Human breast cance
17	6287	92.3	1200	21	AAB21208	Human HER-2/neu pr
18	5821.5	85.4	1256	21	AAB21199	Rat HER-2/neu prot
19	5821.5	85.4	1256	23	AAM51144	Rat Her-2/neu onco
20	5794.5	85.0	1256	21	AAB21206	Mouse Her-2/neu pr
21	5794.5	85.0	1256	22	AAG62860	Amino acid sequenc
22	5794.5	85.0	1256	23	AAM51151	Mouse Her-2/neu on
23	4815	70.7	919	21	AAB21203	Human HER-2/neu fu
24	4815	70.7	919	23	AAM51148	Her-2/neu extracel
25	4065.5	59.7	920	23	AAM51152	Mouse Her-2/neu ex
26	4065.5	59.7	926	23	AAM51153	Mouse Her-2/neu ex
27	3699	54.3	712	21	AAB21204	Human HER-2/neu fu
28	3699	54.3	712	23	AAM51149	Her-2/neu extracel
29	3552	52.1	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	52.1	653	21	AAB21200	Extracellular HER-
31	3550	52.1	653	23	AAM51145	Human Her-2/neu on
32	3512	51.5	645	22	AAB60408	Human ErbB2 oncopr
33	3512	51.5	645	22	AAB61593	Human ErbB2 extrac
34	3447	50.6	951	21	AAY44993	DC8scFv-erbB2EC fu
35	3344	49.1	624	11	AAR08222	Extracellular port
36	3069	45.0	1210	21	AAB19259	Amino acid sequenc
37	3069	45.0	1210	21	AAY50616	Human EGF receptor
38	3069	45.0	1210	23	AAE23019	Human Her-1 protei
39	3069	45.0	1210	23	AAM50768	Human epidermal gr
40	3067	45.0	1210	22	AAB68420	Amino acid sequenc
41	3053.5	44.8	654	21	AAB21205	Rat HER-2/neu prot
42	3053.5	44.8	654	23	AAM51150	Rat Her-2/neu onco
43	3028	44.4	1210	23	ABP51768	Human epidermal gr
44	2977	43.7	583	23	AAE20483	Human protein for
45	2977	43.7	587	23	AAE20481	Human protein for

ALIGNMENTS

RESULT 1
AAY92620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= insertion_region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

```
FT                   /label= Cysteine_rich_domain
FT   Region          210..224
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          250..264
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          324..483
FT                   /label= Ligand_binding_domain
FT   Region          325..339
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          369..383
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          465..479
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          484..623
FT                   /label= Cysteine_rich_domain
FT   Region          579..593
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          624..654
FT                   /label= Transmembrane_domain
FT   Region          632..652
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          653..667
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          655..1010
FT                   /label= Tyrosine_kinase_domain
FT   Region          661..675
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          695..709
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          710..730
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          1011..1235
FT                   /label= C-terminal_domain
XX
PN   WO200020027-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;    99WO-DK00525.
XX
PR   05-OCT-1998;    98DK-0001261.
PR   20-OCT-1998;    98US-0105011.
XX
PA   (MEBI-) M & E BIOTECH AS.
XX
PI   Steinaa L,  Mouritsen S,  Nielsen KG,  Haaning J,  Leach D,  Dalum I;
PI   Gautam A,  Birk P,  Karlsson G;
XX
DR   WPI; 2000-349917/30.
DR   N-PSDB; AAA09455.
XX
PT   Inducing immune responses to weakly immunogenic, tumor associated
PT   peptide antigens for the treatment of breast and prostate cancer
XX
PS   Claim 62; Page 193-198; 220pp; English.
XX
CC   This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC   Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC   response.  Subdominant CTL epitopes, antibody binding regions and
CC   cysteine residues involved in disulfide bonds are preserved in the
CC   immunogenized forms.  Regions suitable for the insertion of foreign T
CC   helper epitopes were identified (see features table).  The method
CC   is used for inducing immune responses against weakly immunogenic
CC   cell-associated peptide antigens (PA) such as those associated with
CC   cancers (self-proteins), e.g. human prostate specific membrane antigen
CC   (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC   The method comprises effecting simultaneous presentation by antigen
CC   producing cells (APCs) of the animals immune system of: (1) at least 1
CC   CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC   B-cell group derived from the cell-associated PA; and (2) at least 1
CC   first T helper cell group which is foreign to the animal.  Analogues of
CC   human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC   part of all known and predicted CTL and B-cell epitopes of the respective
CC   PA and including at least one foreign T helper epitope are also claimed.
CC   The method is used to treat prostate, prostate/breast or breast cancer
CC   when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ   Sequence   1255 AA;

  Query Match          97.0%;  Score 6610;  DB 21;  Length 1255;
  Best Local Similarity   96.9%;  Pred. No. 0;
  Matches 1220;  Conservative   10;  Mismatches   19;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNPTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:   |:  ::    :|||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776
```

```
Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 2
AAE12130
ID   AAE12130 standard; Protein; 1255 AA.
XX
AC   AAE12130;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human tyrosine kinase-type receptor, HER-2.
XX
KW   Therapeutic compound; major histocompatibility complex; vaccine;
KW   antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW   adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW   antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          774..782
FT                   /note= "Antigenic epitope"
XX
PN   WO200168677-A2.
XX
PD   20-SEP-2001.
XX
PF   16-MAR-2001; 2001WO-US40328.
XX
PR   16-MAR-2000; 2000US-0527487.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Nicolette CA;
XX
DR   WPI; 2001-616284/71.
DR   N-PSDB; AAD19731.
XX
PT   Novel synthetic therapeutic compound for inducing immune response and
PT   for use in adoptive immunotherapy, has enhanced binding to major
PT   histocompatibility molecules and enhanced immunoregulatory properties
PT   -
XX
PS   Claim 4; Page 63-67; 69pp; English.
XX
CC   The invention relates to synthetic therapeutic compounds (antigenic
CC   peptides) with enhanced binding to major histocompatibility complex
CC   (MHC) molecules and enhanced immunoregulatory properties relative
CC   to their natural counterparts. Compounds of the invention are useful
CC   for inducing an immune response in a subject and for use in adoptive
CC   immunotherapy. They are useful as components of anti-cancer vaccines
CC   and to expand immune effector cells that are specific for cancers
CC   characterised by expression of the breast cancer antigen, HER-2.
CC   Polynucleotides that encode peptides of the invention are useful as
CC   hybridisation probes and as primers for the detection of genes of gene
CC   transcripts that are expressed in antigen presenting cells (APCs), to
CC   confirm transduction of polynucleotides into host cells. The present
CC   sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC   of the invention are designed based on the HER-2 antigenic peptide
CC   (774-782).
XX
SQ   Sequence   1255 AA;

  Query Match             97.0%;  Score 6610;  DB 22;  Length 1255;
  Best Local Similarity   96.9%;  Pred. No. 0;
  Matches 1220;  Conservative   10;  Mismatches   19;  Indels   10;  Gaps    2;
```

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||         :|:   |: ::    :||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716
```

```
Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX
DR WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;

Query Match 97.0%; Score 6610; DB 22; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEPAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEPAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:  |: ::   :|||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 4
AAU74545
ID   AAU74545 standard; Protein; 1255 AA.
XX
AC   AAU74545;
XX
DT   23-APR-2002  (first entry)
XX
DE   Human HER2 (ErbB2) polypeptide.
XX
KW   Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW   anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW   stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW   thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW   glial disorder; astrocytal disorder; hypothalamic disorder;
KW   glandular disorder; macrophagal disorder; epithelial disorder;
KW   stromal disorder; blastocoelic disorder; inflammatory disorder;
KW   angiogenic disorder; immunological disorder.
XX
OS   Homo sapiens.
XX
PN   US2002001587-A1.
XX
PD   03-JAN-2002.
XX
PF   16-MAR-2001; 2001US-0811123.
XX
PR   16-MAR-2000; 2000US-189844P.
PR   05-OCT-2000; 2000US-238327P.
XX
PA   (ERIC/) ERICKSON S.
PA   (SCHW/) SCHWALL R.
PA   (SLIW/) SLIWKOWSKI M.
XX
PI   Erickson S,  Schwall R,  Sliwkowski M;
XX
DR   WPI; 2002-163686/21.
DR   N-PSDB; ABK14058.
XX
PT   Treating tumour characterised by overexpression of epidermal growth
PT   factor receptor, ErbB or cancer in mammal, comprises administering
PT   anti-ErbB antibody-maytansinoid conjugate to the mammal  -
XX
PS   Example 3; Fig 7; 93pp; English.
XX
CC   The invention relates to treating a tumour in a mammal, where the tumour
CC   is characterised by the overexpression of an epidermal growth factor
CC   receptor (ErbB) and does not respond or responds poorly, to treatment
CC   with an anti-ErbB antibody, comprising administering to the mammal an
CC   anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC   treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC   salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC   prostate and bladder, preferably breast cancer. The breast cancer is a
CC   metastatic breast cancer or an aggressive form of metastatic breast
CC   cancer which overexpresses ErbB2. The method is also useful for treating
CC   neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC   epithelial, stromal, blastocoelic, inflammatory, angiogenic and
CC   immunological disorders. This sequence represents the human HER2 (ErbB2)
CC   polypeptide of the invention.
XX
SQ   Sequence   1255 AA;

  Query Match             97.0%;  Score 6610;  DB 23;  Length 1255;
  Best Local Similarity   96.9%;  Pred. No. 0;
  Matches 1220;  Conservative   10;  Mismatches   19;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:   |: ::    :|||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 5
AAW01111
ID   AAW01111 standard; Protein; 1255 AA.
XX
AC   AAW01111;
XX
DT   01-JAN-1997  (first entry)
XX
DE   HER-2/neu protein.
XX
KW   HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW   breast cancer; ovary cancer; colon cancer; lung cancer;
KW   prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          676..1255
FT                   /label= Intracellular_domain
FT                   /note= "claimed domain, useful for immunisation"
XX
PN   WO9630514-A1.
XX
PD   03-OCT-1996.
XX
PF   28-MAR-1996;   96WO-US01689.
XX
PR   31-MAR-1995;   95US-0414417.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Disis ML;
XX
DR   WPI; 1996-455361/45.
DR   N-PSDB; AAT40739.
XX
PT   DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT   treatment of malignancies with which the HER-2/neu oncogene is
PT   associated
XX
PS   Claim 2; Page 56-61; 71pp; English.
XX
CC   Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC   the product of the HER-2/neo oncogene (see also AAT40739).  The
CC   protein is over-expressed in various cancers, including breast,
CC   ovarian, colon, lung and prostate.  The intracellular domain of the
CC   protein can be used to immunise an animal against a malignancy with
CC   which the oncogene is associated.  The polypeptide can be produced
CC   in transformed host cells for use in immunisation.  Alternatively,
CC   animal cells are transfected in vivo or ex vivo with a viral vector
CC   that directs expression of the polypeptide.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6604;  DB 17;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||         :|:   |:  ::    :||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836
```

```
Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 6
AAW92406
ID   AAW92406 standard; Protein; 1255 AA.
XX
AC   AAW92406;
XX
DT   21-APR-1999  (first entry)
XX
DE   Human HER-2/neu oncogene protein.
XX
KW   HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW   malignancy; treatment; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          676..1255
FT                   /note= "region which elicits immune response"
XX
PN   US5869445-A.
XX
PD   09-FEB-1999.
XX
PF   01-APR-1996;   96US-0625101.
XX
PR   01-APR-1996;   96US-0625101.
PR   17-MAR-1993;   93US-0033644.
PR   12-AUG-1993;   93US-0106112.
PR   31-MAR-1995;   95US-0414417.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Disis ML;
XX
DR   WPI; 1999-152835/13.
DR   N-PSDB; AAX01912.
XX
PT   Use of HER-2/neu polypeptides - for eliciting an immune response to
PT   an HER-2/neu associated malignancy, particularly for treating or
PT   preventing tumours
XX
PS   Claim 3; Column 31-38; 26pp; English.
XX
CC   This sequence represents the human HER-2/neu oncogene protein. A fragment
CC   of this protein is used in a method for eliciting or enhancing an immune
CC   response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC   B cells to produce an immune response to the HER-2/neu protein. The
CC   method can be used for immunisation against a malignancy in which the
CC   HER-2/neu oncogene is associated and in the treatment of an existing
CC   tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match             96.9%;  Score 6604;  DB 20;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:   |:  ::    :||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896
```

```
Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 7
AAB21198
ID   AAB21198 standard; protein; 1255 AA.
XX
AC   AAB21198;
XX
DT   12-JAN-2001  (first entry)
XX
DE   Human HER-2/neu protein.
XX
KW   Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW   breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW   colon cancer.
XX
OS   Homo sapiens.
XX
PN   WO200044899-A1.
XX
PD   03-AUG-2000.
XX
PF   28-JAN-2000; 2000WO-US02164.
XX
PR   29-JAN-1999;   99US-0117976.
XX
PA   (CORI-) CORIXA CORP.
PA   (SMIK ) SMITHKLINE BEECHAM.
XX
PI   Cheever MA,  Gheysen D;
XX
DR   WPI; 2000-505976/45.
DR   N-PSDB; AAA89736.
XX
PT   HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT   useful for vaccinating against breast, ovarian, colon, lung and
PT   prostate cancers -
XX
PS   Claim 52; Fig 7; 128pp; English.
XX
CC   The present sequence is the human HER-2/neu protein. It is a member of
CC   the tyrosine kinase family of receptor-like glycoproteins and shows
CC   homology to the epidermal growth factor receptor (EGFR). It probably
CC   plays a part in cell growth and/or differentiation. The HER-2/neu
CC   gene is an oncogene. An HER-2/neu fusion protein comprising a
CC   HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC   domain may be used to treat or prevent cancer by eliciting or
CC   enhancing an immune response to the HER-2/neu protein. It may be used
CC   to treat malignancies such as breast, ovarian, colon, lung and
CC   prostate cancers, and may be used as an antigen to vaccinate against
CC   these neoplasias.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6604;  DB 21;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:   |: ::   :||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
```

```
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074

Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134

Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194

Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253

Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
PN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 21; Length 1255;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714

Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774

Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834

Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894

Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954

Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014

Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
```

```
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134

Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194

Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253

Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 9
AAB85458
ID   AAB85458 standard; Protein; 1255 AA.
XX
AC   AAB85458;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human HER-2/neu protein.
XX
KW   Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW   oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS   Homo sapiens.
XX
PN   WO200153463-A2.
XX
PD   26-JUL-2001.
XX
PF   19-JAN-2001; 2001WO-US01850.
XX
PR   21-JAN-2000; 2000US-0177545.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Cheever MA,  Hand-Zimmermann S;
XX
DR   WPI; 2001-476112/51.
DR   N-PSDB; AAH23392.
XX
PT   New antigen-presenting cells, useful as vaccines for eliciting or
PT   enhancing an immune response to HER-2/neu protein, particularly useful
PT   for treating or preventing cancer, e.g. breast cancer  -
XX
PS   Claim 2; Page 41-46; 49pp; English.
XX
CC   The invention provides an isolated antigen-presenting cell, which
CC   expresses at least an immunogenic portion of a polypeptide that produces
CC   an immune response to HER-2/neu protein. The antigen-presenting cells are
CC   useful as vaccines for eliciting or enhancing an immune response to
CC   HER-2/neu protein, particularly in treating or preventing malignancies in
CC   which the HER-2/neu oncogene is associated. Specifically, these are
CC   useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC   colon, lung or prostate cancers. The present sequence represents
CC   the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ   Sequence   1255 AA;

  Query Match           96.9%;  Score 6604;  DB 22;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714

Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774

Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834

Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894

Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954

Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014

Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074

Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134

Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
```

```
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 10
AAG88267
ID   AAG88267 standard; Protein; 1255 AA.
XX
AC   AAG88267;
XX
DT   11-SEP-2001  (first entry)
XX
DE   HER2/neu amino acid sequence.
XX
KW   Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW   immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW   tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS   Homo sapiens.
XX
PN   WO200141787-A1.
XX
PD   14-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US33591.
XX
PR   10-DEC-1999;   99US-0458299.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
PI   Keogh E;
XX
DR   WPI; 2001-374995/39.
XX
PT   An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT   cellular immune responses for the prevention and treatment of cancer -
XX
PS   Disclosure; Page 15; 199pp; English.
XX
CC   The present invention describes isolated prepared HER2/neu epitopes (I).
CC   Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC   culture in vitro and binds to a complex of an epitope (I), bound to a
CC   human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC   and a second epitope and the peptide is less than 50 contiguous amino
CC   acids that have 100% identity with a native peptide sequence of HER2/neu;
CC   (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC   excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC   (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC   and immunostimulant activities, and can be used in vaccines. (I), (II)
CC   and (III) are useful for inducing cellular immune responses for the
CC   prevention and treatment of cancer. (I) and (II) are useful for
CC   monitoring or evaluating an immune response to a tumour-associated
CC   antigen when incubated with a T lymphocyte sample form a patient and
CC   detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC   based vaccines mean that immunosuppressive epitopes that may be present
CC   in whole antigens may be avoided. Selected epitopes may be combined to
CC   enhance immunogenicity. The possible pathological side effects caused by
CC   infectious agents or whole protein antigen is eliminated. The vaccine
CC   provides the ability to direct and focus an immune response to multiple
CC   selected antigens from the same pathogen. Epitope-based anti-tumour
CC   vaccines provides the opportunity to combine epitopes derived from
CC   multiple tumour-associated molecules addressing the problem of tumour-
CC   tumour variability and reducing the likelihood of tumour escape due to
CC   antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6604;  DB 22;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy         1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy        61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy       121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy       181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy       361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy       661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
             |||||||||||||||      :|:  |: ::   :||||||||||||||||||||||||
Db       661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy       715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy       775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy       835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy       895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy       955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy      1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 11
AAE24067
ID   AAE24067 standard; Protein; 1255 AA.
XX
AC   AAE24067;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Human Her-2 protein.
XX
KW   Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW   hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW   tumour; gene therapy; phosphorothioate backbone.
XX
OS   Homo sapiens.
XX
PN   WO200222636-A1.
XX
PD   21-MAR-2002.
XX
PF   12-SEP-2001; 2001WO-US28572.
XX
PR   15-SEP-2000; 2000US-0663834.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Cowsert LM;
XX
DR   WPI; 2002-471192/50.
DR   N-PSDB; AAD38904.
XX
PT   Novel antisense oligonucleotide which modulates the expression of Human
PT   Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT   inflammation or to prevent infection in humans -
XX
PS   Example 13; Page 95-107; 116pp; English.
XX
CC   The invention relates to antisense compounds targetted to a nucleic
CC   acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC   that specifically hybridises with and inhibits the expression of Her2.
CC   Antisense compounds of the invention are used for treating diseases or
CC   conditions associated with Her2 such as hyperproliferative disorders
CC   e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC   neural or cardiac cancer. They are also useful prophylactically e.g.
CC   to prevent or delay infection, inflammation and tumour formation. The
CC   invention is also used in gene therapy. The present sequence is human
CC   Her-2 protein.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6604;  DB 23;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:   |: ::    :||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
```

```
                  |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 12
AAE20479
ID   AAE20479 standard; Protein; 1255 AA.
XX
AC   AAE20479;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Human Her-2/neu protein.
XX
KW   Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW   human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1021..1030
FT                   /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN   WO200214503-A2.
XX
PD   21-FEB-2002.
XX
PF   14-AUG-2001; 2001WO-US41733.
XX
PR   14-AUG-2000; 2000US-225152P.
PR   28-SEP-2000; 2000US-236428P.
PR   21-FEB-2001; 2001US-270520P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Hand-zimmermann S,  Cheever MA,  Foy TM,  Lodes MJ,  Kalos MD;
PI   Mcneill PD,  Vedvick TS;
XX
DR   WPI; 2002-280758/32.
DR   N-PSDB; AAD32743.
XX
PT   Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT   prevention and diagnosis of cancer, preferably breast cancer  -
XX
PS   Disclosure; Page 114-117; 129pp; English.
XX
CC   The invention relates to an isolated Her-2/Neu polypeptide composition
CC   effective for eliciting an immune response. The invention is useful for
CC   eliciting an immune response in a patient, where the patient is human
CC   leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC   The composition is useful for the therapy and diagnosis of cancer,
CC   preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC   and other compositions for the diagnosis, prevention and treatment of
CC   human malignancies, for stimulating and/or expanding T cells specific for
CC   Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC   patient. The invention is useful for stimulating a T cell response in a
CC   human patient, as probe or primer for nucleic acid hybridisation, to
CC   selectively form duplex molecules with complementary stretches of the
CC   entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC   length gene from a suitable library, and to direct expression of a
CC   polypeptide in appropriate host cells. The composition is useful in
CC   prophylactic or therapeutic applications and for the treatment of cancer,
CC   preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC   associated malignancies. The invention is useful in gene therapy. The
CC   present sequence is human Her-2/neu protein.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6604;  DB 23;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:  |: ::   :||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 13
AAM51143
ID   AAM51143 standard; Protein; 1255 AA.
XX
AC   AAM51143;
XX
DT   17-JUN-2002  (first entry)
XX
DE   Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX
KW   Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW   tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..653
FT                   /note= "extracellular domain"
FT   Domain          676..1255
FT                   /note= "intracellular domain"
FT   Domain          990..1255
FT                   /note= "phosphorylation domain"
XX
PN   WO200212341-A2.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-US24283.
XX
PR   03-AUG-2000; 2000US-0632507.
XX
PA   (CORI-) CORIXA CORP.
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Cheever MA,  Gheysen D;
XX
DR   WPI; 2002-241743/29.
DR   N-PSDB; ABA92250.
XX
PT   Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT   or enhancing an immune response to the protein, has Her-2/neu
PT   extracellular domain fused to Her-2/neu intracellular or
PT   phosphorylation domain  -
XX
PS   Claim 68; Fig 7; 141pp; English.
XX
CC   The present sequence is that of human Her-2/neu (p185 glycoprotein
CC   or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC   vaccines.  The Her-2/neu gene is amplified and p185 is overexpressed
CC   in a variety of cancers, including breast, ovarian, colon, lung and
CC   prostate cancer.  Her-2/neu is a member of the tyrosine kinase
CC   family of receptor-like glycoproteins.  It comprises an extracellular
CC   domain with homology to the epidermal growth factor receptor
CC   (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC   intracellular domain that also shows homology to EGFR.  Its
CC   overexpression correlates with a poor prognosis in breast and
CC   ovarian cancers.  The invention provides Her-2/neu fusion
CC   proteins, nucleic acids encoding them, viral vectors, and vaccines
CC   comprising the fusion proteins or nucleic acid molecules.  In
CC   preferred fusion proteins, the extracellular domain of a Her-2/neu
CC   protein is fused to a Her-2/neu intracellular domain or
CC   phosphorylation domain (or its DeltaPD fragment).  An immune
CC   response to Her-2/neu protein is elicited or enhanced by
CC   administering the fusion protein in the form of a vaccine, or by
CC   transfecting cells of an animal ex vivo with a nucleic acid
CC   encoding the fusion protein, and delivering the transfected cells
CC   to the animal.  The fusion proteins, nucleic acids, and isolated
CC   specific T-cells are useful for inhibiting the development of a
CC   cancer, especially breast, ovarian, colon, lung or prostate cancer
CC   in a patient.  T cells that specifically react with a Her-2/neu
CC   fusion protein can be used to remove tumour cells from a sample in
CC   order to inhibit the development of cancer in a patient.
XX
SQ   Sequence   1255 AA;

  Query Match             96.9%;  Score 6604;  DB 23;  Length 1255;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 1218;  Conservative   11;  Mismatches   20;  Indels   10;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||        :|:   |: ::    :|||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776
```

```
Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide -
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 23; Length 1255;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714
              |||||||||||||||      :|:   |:  ::    :||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy        715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy        775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy        835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894
```

```
Db        837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy        895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954

Db        897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy        955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014

Db        957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy       1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074

Db       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076

Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134

Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194

Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253

Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells. such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isolelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 96.3%; Score 6561; DB 14; Length 1433;
Best Local Similarity 96.2%; Pred. No. 0;

Matches 1211; Conservative 13; Mismatches 25; Indels 10; Gaps 2;

```
Qy           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQVCQVVQGNL 60

Qy          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEHCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy         361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db         361 IQEFAGCKRIFGSLAFLPESFDGDPASNTAPLQPEHLQVFETLEQITGYLYISAWPDSLP 420

Qy         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLSFVHTV 480

Qy         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db         541 VEECRVLQGLPREYVNASHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy         601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Db         601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy         661 ILLVVVLGVVFGILI------FNNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAQMRILK 714

Db         661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL----VEPLTPSGAMPNQAQMRILK 716

Qy         715 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 774

Db         717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAG 776

Qy         775 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 834

Db         777 VGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYL 836

Qy         835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 894

Db         837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILR 896

Qy         895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 954

Db         897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMV 956

Qy         955 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1014

Db         957 KCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGD 1016

Qy        1015 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1074

Db        1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076
```

```
Qy       1075 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136

Qy       1135 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1194
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEY 1196

Qy       1195 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 LTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

Search completed: July 22, 2003, 09:16:38
Job time : 41.875 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.3921 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-695-709-14
Perfect score: 6853
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6641	96.9	1255	1	A24571	protein-tyrosine k
2	5844	85.3	1260	1	TVRTNU	protein-tyrosine k
3	5834.5	85.1	1254	2	I48161	p-185 precursor -
4	3065	44.7	1210	1	GQHUE	epidermal growth f
5	3036	44.3	1210	2	A53183	epidermal growth f
6	3013.5	44.0	1223	1	TVCHLV	epidermal growth f
7	2884.5	42.1	1308	2	A47253	epidermal growth f
8	2592	37.8	1166	1	S06142	protein-tyrosine k
9	2334.5	34.1	1342	2	A36223	kinase-related tra
10	2248.5	32.8	1339	2	JC4387	epidermal growth f
11	1685.5	24.6	698	1	TVFVLV	protein-tyrosine k
12	1622	23.7	604	1	TVYUH	protein-tyrosine k
13	1569	22.9	544	2	S35745	protein-tyrosine k
14	1562	22.8	545	2	S00727	kinase-related tra
15	1557.5	22.7	1330	1	GQFFE	epidermal growth f
16	1545	22.5	540	2	B44776	protein-tyrosine k
17	1543	22.5	540	1	TVFVEB	protein-tyrosine k
18	1509	22.0	644	2	A36325	epidermal growth f
19	1241	18.1	1323	2	E88257	protein let-23 [im
20	1241	18.1	1374	2	S70712	protein-tyrosine k
21	1153	16.8	1369	2	S70713	protein-tyrosine k
22	1129	16.5	1717	1	A45558	epidermal growth f
23	1126	16.4	527	2	A42032	epidermal growth f
24	975.5	14.2	843	2	A27131	epidermal growth f
25	806.5	11.8	346	2	S13807	protein-tyrosine k
26	754.5	11.0	311	2	S13808	protein-tyrosine k
27	700	10.2	1363	2	T43220	insulin-like growt
28	670	9.8	1383	2	A36080	insulin receptor p
29	669.5	9.8	1372	2	A34157	insulin receptor p
30	668	9.7	1382	1	INHUR	insulin receptor p
31	655	9.6	1477	2	T18534	protein-tyrosine k
32	653	9.5	1607	2	T43212	insulin-like growt
33	650.5	9.5	1300	2	A36502	insulin receptor-r
34	638	9.3	1268	2	B36502	insulin receptor-r
35	596	8.7	1367	1	IGHUR1	insulin-like growt
36	590	8.6	1390	2	T30346	insulin receptor -
37	588.5	8.6	2148	1	A56081	insulin receptor -
38	582	8.5	2101	2	S57245	insulin receptor (
39	577	8.4	1371	2	A33837	insulin-like growt
40	574	8.4	1114	1	S05582	protein-tyrosine k
41	571	8.3	987	2	A54092	protein-tyrosine k
42	569.5	8.3	952	2	I50612	protein-tyrosine k
43	566.5	8.3	984	2	A39753	protein-tyrosine k
44	562.5	8.2	977	2	S49004	tyrosine kinase Mp
45	556.5	8.1	976	2	A36355	protein-tyrosine k

ALIGNMENTS

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito,
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal gro
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/ep
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seebu
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares ch
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517,'RALL',522,'S',524-654,'V',656-1169,'A',1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:g183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcript
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

```
A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TMM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

  Query Match             96.9%;  Score 6641;  DB 1;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 5.5e-266;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||

Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||  |    ||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255


RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related prote
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Co
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
```

```
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

  Query Match             85.3%;  Score 5844;  DB 1;  Length 1260;
  Best Local Similarity   85.7%;  Pred. No. 3.2e-233;
  Matches 1083;  Conservative   52;  Mismatches  119;  Indels   10;  Gaps    4;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||||
Db          4 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 63

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| |||||||||||||||:|||||||::|||||||||||||||| ||||||||
Db         64 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 123

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||||: |||||||| :||||:| |||||
Db        124 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 183

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |   |||||||| ||:| || : ||||| |||| || |:|  |||||| ||||||||
Db        184 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 243

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||| |
Db        244 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 303

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||:::::
Db        304 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 363

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
                 || |||||||||||||||||:|  |||:|||||||||||||||||||||||||||||
Db        364 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 423

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              |||||||::|||||||:|||||||||||||  |||||||||||||| | |||||||
Db        424 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 483

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||: |||:: ||  || |: ||| ||||||||||||| |||||
Db        484 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 543

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : |||||||||||| | |||| ||||| |||||||    |||
Db        544 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 603

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAV 658
              |||||||||||||||||:||||| |||||||||||||||::||||||||||:| |:: |
Db        604 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRASPVTFIIATV 663

Qy        659 VGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEF 718
              |:|| ::|  || |||||||:|||||||||||||||||||||||||||||||||||||||
Db        664 EGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 723

Qy        719 NNFTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 777
                        ||  ||  |  |||||||||||||||||||||||||||||||||||||||
Db        724 E-------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 776

Qy        778 YVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAK 837
              |||||||||||||||||||||||||||||||||||||||:|||||||||||||:||||
Db        777 YVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAK 836

Qy        838 GMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMAL 897
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 GMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMAL 896

Qy        898 ESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDV 957
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 ESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDV 956

Qy        958 YMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLED 1017
              |||||||||||||||||||||||||||||||||||||||||||||:||:||||||||||
Db        957 YMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLED 1016

Qy       1018 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAP 1077
              |||||||||||||||||| ||| || |   | ||||||||||:|||||||||  |
Db       1017 DDMGDLVDAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPP 1076

Qy       1078 RSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPL 1137
              |||||||||||||||||| ||  ||||||  || ||||||||||:||| |||||||||
Db       1077 RSPLAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPL 1136

Qy       1138 TCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAV 1197
              |||||||||| :|:|||||   |||||  |||||||||| ||||||||||||||||||||
Db       1137 ACSPQPEYVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAV 1196

Qy       1198 ENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGL 1257
              ||||||  |:  | |   ||| |||||||||||||||:   |:| ||| |:||||||||||
Db       1197 ENPEYLVPREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGL 1256

Qy       1258 DVPV 1261
              ||||
Db       1257 DVPV 1260


RESULT 3
I48161
p-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishi
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MUID:94193007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

  Query Match             85.1%;  Score 5834.5;  DB 2;  Length 1254;
  Best Local Similarity   85.3%;  Pred. No. 7.8e-233;
  Matches 1077;  Conservative   61;  Mismatches  115;  Indels    9;  Gaps    3;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| | |||||||| |||: |||||||||||||||||||||||::||||||||||||
Db          1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:||||:||| |||||||||||||||| ||||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||:| |  || :| |||||||||||||||||||: |||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||| ||:| || : ||| | ||||:|| ||:         :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240
```

```
Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||:|||||||||||:|||||||||||||||||| ||||||||||::::
Db        301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||:|:|   ||| ||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||:|||||:||:|||||  ||||||  |||||||||||||||  ||| ||||||||||
Db        421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||: |  |:||  :  ||: ||| ||||||||||||||| |||||||
Db        481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |:|||| :|||||||| :|||||||||||||| : || | |||| || |||| |||||||
Db        541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              |||||||||||||||:|||| ||||||||||||||::|||||||||| |||:: |||
Db        601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||| :|:||| |||||||:|||||||||||||||||||||||||||||||||||||||||
Db        661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                    ||  ||   |    ||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              |||:||||||||||||||||||||||||||||||:|||||||||||||:||||||
Db        774 MAGLGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              |||||||||||||||| ||||||||   | ||||||||||:||||:||| || |||
Db       1014 MGDLVDAEEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              |||||||||||||:|:||||| || ||:   | ||||||||||:|||:|||||||| |
Db       1074 PLAPSEGAGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLAC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||:||||||   |||||  |||||||| ||||||||||||||  |||||||
Db       1134 SPQPEYVNQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              |||| |:||:| ||| |||  ||||||||||||| |||:||:||:|||||||||||||||
Db       1194 PEYLVPRGGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDV 1252

Qy       1260 PV 1261
              ||
Db       1253 PV 1254
```

RESULT 4
GQHUE
epidermal growth factor receptor precursor - human
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Homo sapiens (man)
C;Date: 15-Nov-1984 #sequence_revision 27-Nov-1985 #text_change 11-Jun-1999
C;Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143
R;Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee,
rg, P.H.
Nature 309, 418-425, 1984
A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expressi
A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210 <ULL>
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epiderma
A;Reference number: A25772; MUID:85270438; PMID:2991899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identificat
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript ter
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 'RCAWRRA',150-187,'KSVIQAV',195,'M',197,'A',199-222,'S',224-304,'RA',30
',798-799,'TD',802-811,'R',813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF
R;Lin, C.R.; Chen, W.S.; Kruiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verm
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30,'S',32-51;454-467 <WEB>

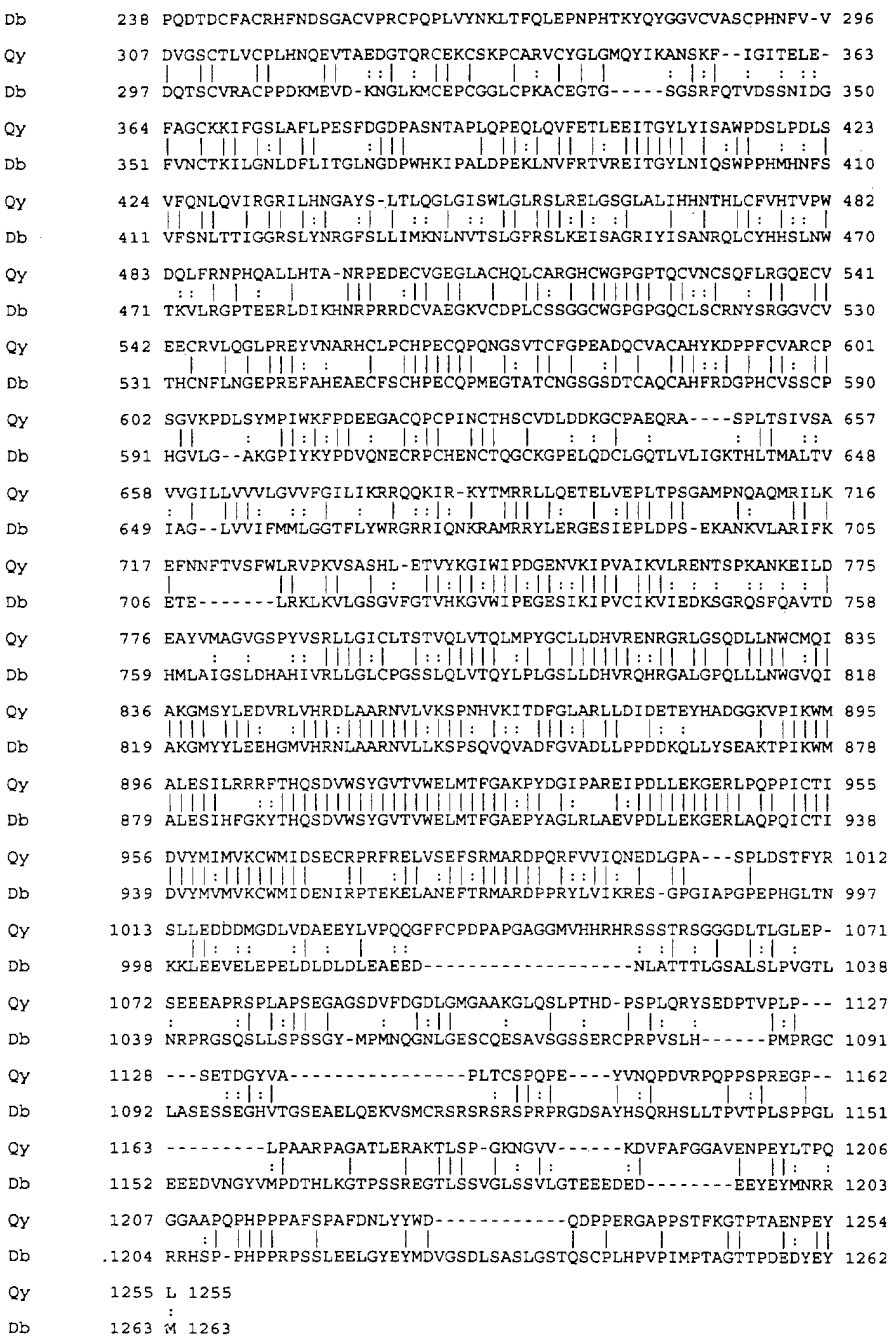
```
R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A;Title: Identification of residues in the nucleotide binding site of the epidermal grow
A;Reference number: A60143; MUID:85182650; PMID:2985580
A;Accession: A60143
A;Molecule type: protein
A;Residues: 740-744,'X',746-747 <RUS>
R;Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and supercc
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity
A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain r
A;Reference number: A33331; MUID:90003233; PMID:2790960
A;Contents: annotation; internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:120610; OMIM:131550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1210/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <EE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <EE2>
F;646-668/Domain: transmembrane #status predicted <TMM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-726/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F;745/Active site: Lys #status experimental

  Query Match             44.7%;  Score 3065;  DB 1;  Length 1210;
  Best Local Similarity   48.7%;  Pred. No. 5.4e-119;
  Matches  620;  Conservative  179;  Mismatches  359;  Indels  114;  Gaps   24;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
Db        127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy        189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
Db        184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy        248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
Db        244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy        308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
Db        304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy        367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
Db        362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy        427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486

Db        422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy        487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
Db        482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCKL 541

Qy        547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606
Db        542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy        607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG---ILL 663
Db        602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLLL 658

Qy        664 VVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNNFTV 723
Db        659 VVALGIG---LFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKE------ 709

Qy        724 SFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGV 783
Db        710 TEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASV 769

Qy        784 GSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLE 843
Db        770 DNPHVCRLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLE 829

Qy        844 DVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRR 903
Db        830 DRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHR 889

Qy        904 RFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVK 963
Db        890 IYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVK 949

Qy        964 CWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGD 1022
Db        950 CWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDD 1009

Qy       1023 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1082
Db       1010 VVDADEYLIPQQGFF---------------------------------SSPSTSRTPLL 1035

Qy       1083 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCS 1140
Db       1036 SSLSATSN--NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------ 1087

Qy       1141 PQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENP 1200
Db       1088 PVPEYINQ-SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNP 1136

Qy       1201 EYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTF 1243
Db       1137 EYLNTVQ---------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIF 1187

Qy       1244 KGTPTAENPEYL 1255
Db       1188 KGS-TAENAEYL 1198


RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C;Accession: A53183; A43818; S24942; A28941; S45325; I49643
R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF recept
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:U03425
```

```
R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding sit
A;Reference number: A43818; MUID:91232866; PMID:2030916
A;Accession: A43818
A;Molecule type: mRNA
A;Residues: 1-714 <AVI>
A;Cross-references: GB:X59698
R;Eisinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A;Reference number: S24942
A;Accession: S24942
A;Molecule type: mRNA
A;Residues: 969-971,'K',973-1115,'D' <EIS>
A;Cross-references: EMBL:Z12608
R;Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A;Reference number: A28941; MUID:88330814; PMID:3138233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971,'K',973-1210 <VER>
A;Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A;Reference number: I49643; MUID:93126380; PMID:7678348
A;Accession: I49643
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 12-20,22-132 <RES>
A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
C;Genetics:
A;Gene: EGFR
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphopro
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TMM>
F;712-977/Domain: protein kinase homology <KIN>
F;720-728/Region: protein kinase ATP-binding motif
F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

  Query Match             44.3%;  Score 3036;  DB 2;  Length 1210;
  Best Local Similarity   48.6%;  Pred. No. 8.4e-118;
  Matches  622;  Conservative  172;  Mismatches  367;  Indels  118;  Gaps   26;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 LVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNNF 721
Db        657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKE---- 711

Qy        722 TVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 781
Db        712 --TEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMA 769

Qy        782 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 841
Db        770 SVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNY 829

Qy        842 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 901
Db        830 LEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESIL 889

Qy        902 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 961
Db        890 HRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIM 949

Qy        962 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDM 1020
Db        950 VKCWMIDADSRPKFRELILEFSQMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDM 1009

Qy       1021 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1080
Db       1010 EDVVDADEYLIPQQGFF----------------NSPST-----------------SRTP 1035

Qy       1081 LAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLT 1138
Db       1036 LLSSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL---- 1087

Qy       1139 CSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVE 1198
Db       1088 --PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVG 1134

Qy       1199 NPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPS 1241
Db       1135 NPEYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNG 1185

Qy       1242 TFKGTPTAENPEYLGLDVP 1260
Db       1186 IFKG-PTAENAEYLRVAPP 1203
```

```
RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Gallus gallus (chicken)
C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C;Accession: A27720; A00643
R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A;Reference number: A27720; MUID:88261272; PMID:3260329
A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
pecific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TMM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

  Query Match             44.0%;  Score 3013.5;  DB 1;  Length 1223;
  Best Local Similarity   47.5%;  Pred. No. 7.2e-117;
  Matches  620;  Conservative  177;  Mismatches  354;  Indels  153;  Gaps   28;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
              |   :| ||  | |         :|| ||: ||      | |   |: :|  |:||  |||
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  |:||: |||| |||||| | |  :||: |:|:||  |:::::||||| |
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
              :| |          |||||  :: |:||| ||| |  ||:||  ||:|| ||  : :  |
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              |::|  :   :|  | | |    |||   ::||:|: :||  |: ||:| :|:|||| |
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: ||||||  |      |:  || || ||  |::   ||||:|:|||:||  |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||   |      ||  |:| ::|:||   |::|| |:|:  :|     |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              |:: |  | || | :: || :| ||  : | || |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
              ||  |:||::|||   :| ||| :  | |  ||||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: |  ||     |       || :::|  :   |  ||:  ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
              |||::| :||| |||:     ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIV 655
              ||  ||:||  :   : :||: |     || |  |||  |     :|||      |  ||
Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 SAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRI 714
              : || |:| :||:|:  |: ::||   :|| |:|||||| |||||||||| |||| :||
Db        657 AGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRI 715

Qy        715 LKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIL 774
              |||      : : :|  : :    |||||:|||:|| |||||||| ||| |||||||||
Db        716 LKE------TEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEIL 769

Qy        775 DEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQ 834
              |||||| | :|:| |||||||||||||:|||||||||||::||::  :||| |||||:|
Db        770 DEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWCVQ 829

Qy        835 IAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKW 894
              |||||:|||: |||||||||||||:| ||||||||||:|| || ||||:|||||||
Db        830 IAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPIKW 889

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954
              ||||||| | :|||||||||||||||:|||||||| ||  :|||||||||||||
Db        890 MALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPICT 949

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRS 1013
              |||||||||||||:: ||:||||::|||:|||| |::||| :| :   || || |||:
Db        950 IDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFYRT 1009

Qy       1014 LLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSE 1073
              |:|::|| |:|||:|||| ||||                | ||
Db       1010 LMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-------------- 1038

Qy       1074 EEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPS 1128
                  |:||     | |  : ::  |         :  |  |  : | :|||| |||
Db       1039 ---SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNFLE 1087

Qy       1129 ET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGV 1186
              |:  ||::      | ||||||  : |: ||                        :  :| :
Db       1088 ESIDDGFL------PAPEYVNQ--LMPKKPS--------------------TAMVQNQI 1118

Qy       1187 VKDVF--------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQ 1231
                 ::                :   ||:|||||          :  |      |::  || |
Db       1119 YNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFESSPYWIQ 1170

Qy       1232 --------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1260
                     | |:          |:      | ||||||  :  |
Db       1171 SGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 1214


RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide
```

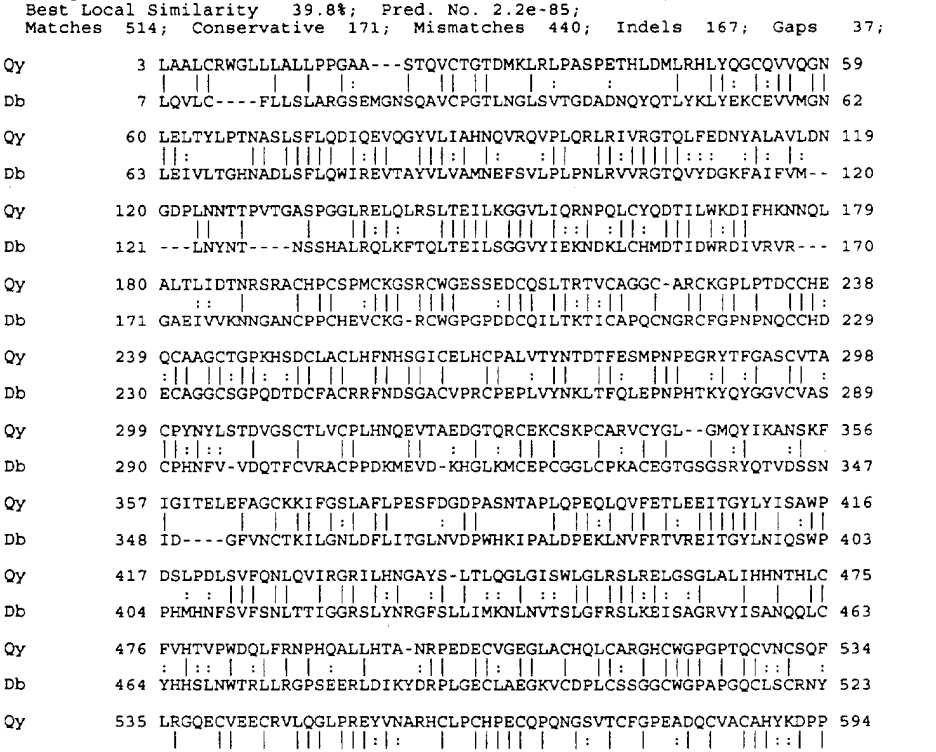
```
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

  Query Match             42.1%;  Score 2884.5;  DB 2;  Length 1308;
  Best Local Similarity   44.2%;  Pred. No. 1.5e-111;
  Matches  598;  Conservative  185;  Mismatches  390;  Indels  179;  Gaps   29;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  |  ||||: ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::||||| ||| : :|  ||| ||| |:||          |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |  | | |||| :  ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |    ||    ||  ||:   |   :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||    ||  |:| : |: |:  | : | |:|   :  :         :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :    : ||:| || |: ||||:| | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||  | ||:|::|    | |:  ||: |  :||:|: :|    |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||   :| ::   ||  : |  ||: |: ||:   ||||| ||::| :| ||: |:|
Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              | :  |  ||:  |   |: | |:|:  ::| :|| ||  |  |:|:|| | || :||
Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PAEQRASPL 651
              |::   |:  |:|: | :  | ||  |||  |          |    | :||
Db        596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQHAR-TPL 652

Qy        652 TSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710
                | : |: |: ::|::|: | : ::|:  | :|  :|| | ||||||||||||  ||||
Db        653 --IAAGVIGGLFILVIVGLTFAVYVRRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNQA 708

Qy        711 QMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKAN 770
              |:|||||      :   ||  : :    |||||||:|:|| |||||||:|  | | ||||
Db        709 QLRILKE------TELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKAN 762

Qy        771 KEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLN 830
              | :||| :|| :  |:: |||:||: |:||||||:||||::|  |::   :||| |||
Db        763 VEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLLN 822

Qy        831 WCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKV 890
                                                                        ||:|||||| |||: |||||||||||||||||||||||||||||||: || ||:|||||:
Db        823 WCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKM 882

Qy        891 PIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQP 950
              ||||||||  |  |:|||||||||||:|||||||  |||||||  |||||||||||||||
Db        883 PIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQP 942

Qy        951 PICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDST 1009
              |||||||||:||||||||:: ||:|:|| :|||||||||||::|||  :|  :    || ||
Db        943 PICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDSK 1002

Qy       1010 FYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGL 1069
              |:::||:::|: |::|||||||| | |  | |      :   | |  | ||      :|
Db       1003 FFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIGH 1050

Qy       1070 EPSEEEAPRS--------------------------PLAP-SEGAGSDVFDGDLGMGA 1100
                |    | |                          | || ::|| :::||     |
Db       1051 SPPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNGT 1110

Qy       1101 AKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRP 1153
                :  :   : |  |||| |||| |       : :||: |:  |: ||:|  :
Db       1111 LRKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE--- 1167

Qy       1154 QPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQP 1213
                  | |  : |                ||| ::       |::||||     |
Db       1168 ------ENPFVSRR--------------KNGDLQ-------ALDNPEYHNASNG----- 1194

Qy       1214 HPPPA--------------------------------FSPAFDNLYYWDQDPPERGA- 1238
                ||                                  ||||  ||:   | |
Db       1195 -PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRSTL 1253

Qy       1239 -PPSTFKGTPT--------------AENPEYL 1255
                |   :  |               |||||||
Db       1254 QHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transfc
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueler, W.; Raulf, F.; Telling, A.; Rc
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueler, W.; Schartl, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphc
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025,'N',1027-1098,'A',1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

  Query Match             37.8%;  Score 2592;  DB 1;  Length 1166;
  Best Local Similarity   44.1%;  Pred. No. 1.5e-99;
```

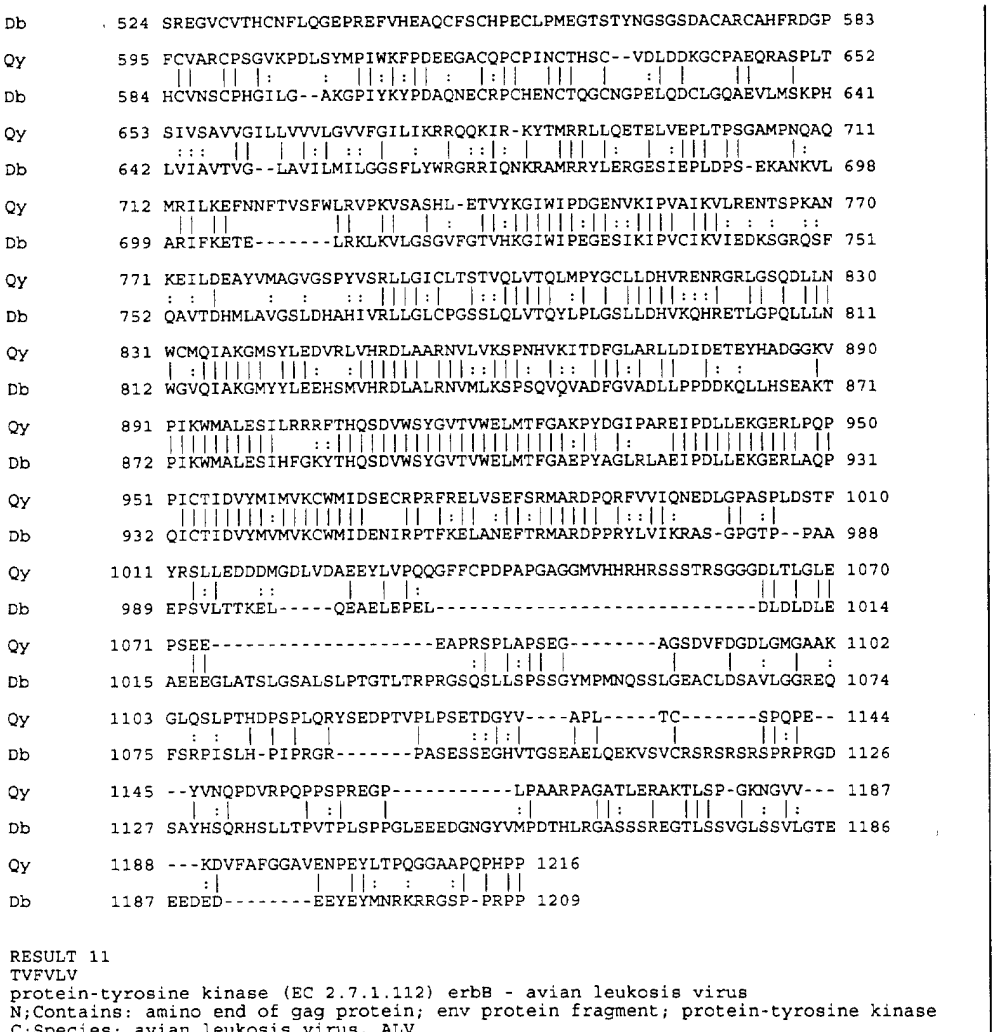
```
  Matches  563;  Conservative  167;  Mismatches  396;  Indels  150;  Gaps   32;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
Db        417 NMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRY 476

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
Db        477 ANTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCW-PGPTMCVSCLHVDR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFC 596
Db        529 GGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKSAHFQDGPQC 588

Qy        597 VARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVS 656
Db        589 IPRCPHGILGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGD-IVSHSSLAVG 646

Qy        657 AVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
Db        647 LVSGLLITVIVALLIVVLLRRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQAFLRILK 705

Qy        717 --EFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIL 774
Db        706 ETEFKKDRV--------LGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQEVL 757

Qy        775 DEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQ 834
Db        758 DEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCVQ 817

Qy        835 IAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKW 894
Db        818 IAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPIKW 877

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954
Db        878 MALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPICT 937

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
Db        938 IEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSLSDRRLFSRL 994

Qy       1015 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1074
Db        995 LSSDD--DVVDADEYLLPYKRI--------------------NRQGS------------ 1019

Qy       1075 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDGY 1133
Db       1020 ----EPCIPPTGH--------------------PVRENSITLRNISDPTQNALEKDLDGH 1055

Qy       1134 VAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKTLS 1180
Db       1056 ----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYLN 1105

Qy       1181 PGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPP 1240
Db       1106 TNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGALT 1140

Qy       1241 STFKGTPTAENPEYLG 1256
Db       1141 GNGMFLPAAENLEYLG 1156


RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C;Accession: A36223; I59164
R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epider
A;Reference number: A36223; MUID:90083234; PMID:2687875
A;Accession: A36223
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1342 <KRA>
A;Cross-references: GB:M29366
R;Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A;Title: Molecular cloning and expression of another epidermal growth factor recept
A;Reference number: I59164; MUID:90311312; PMID:2164210
A;Accession: I59164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-559,'G',561-957,'F',959-1063,'G',1065-1342 <RES>
A;Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C;Genetics:
A;Gene: GDB:ERBB3; HER3
A;Cross-references: GDB:119880; OMIM:190151
A;Map position: 12q13-12q13
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; phosphotransferase
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

  Query Match             34.1%;  Score 2334.5;  DB 2;  Length 1342;
  Best Local Similarity   39.6%;  Pred. No. 6.4e-89;
  Matches  523;  Conservative  195;  Mismatches  460;  Indels  143;  Gaps   36;

Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
```



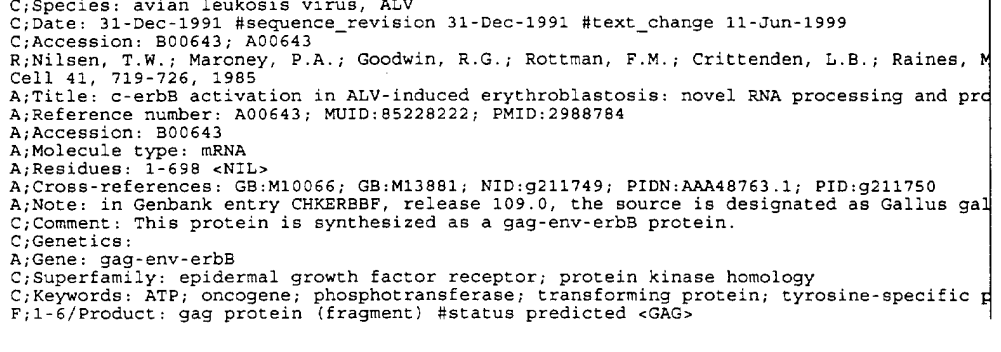
RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant prot
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res
C;Comment: This protein is a functional heregulin receptor that transduces signals
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TMM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Ty

Query Match 32.8%; Score 2248.5; DB 2; Length 1339;
Best Local Similarity 39.8%; Pred. No. 2.2e-85;
Matches 514; Conservative 171; Mismatches 440; Indels 167; Gaps 37;

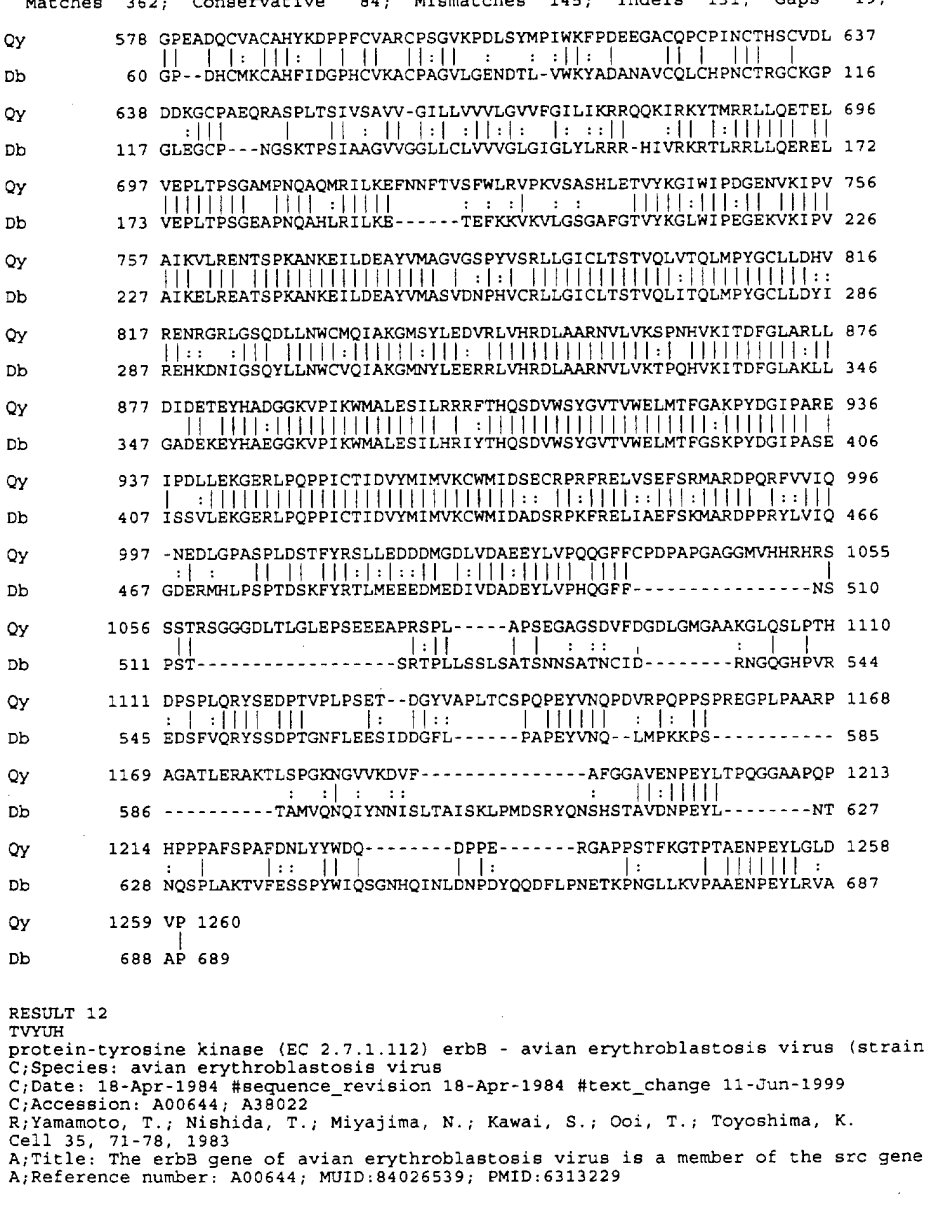




RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and prc
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted



Query Match 24.6%; Score 1685.5; DB 1; Length 698;
Best Local Similarity 50.1%; Pred. No. 1.6e-62;
Matches 362; Conservative 84; Mismatches 145; Indels 131; Gaps 19;



RESULT 12
TVYUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene
A;Reference number: A00644; MUID:84026539; PMID:6313229

```
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

  Query Match             23.7%;  Score 1622;  DB 1;  Length 604;
  Best Local Similarity   50.0%;  Pred. No. 5.8e-60;
  Matches  348;  Conservative   80;  Mismatches  136;  Indels  132;  Gaps   17;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
              |||: | | ||  ||:||  :   : : |: |     || |  |||  |      :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705
                |   || : || |:| :||:|:  |: ::||   :|| |:|||||| ||||||||||
Db           59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENT 765
              |||| :|||||      : : :|  : :    |:|||:|||:|| |||||||| ||| |
Db        118 APNQAHLRILKE------TEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKELREAT 171

Qy        766 SPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGS 825
              |||||||||||||||| | :|:| ||||||||||||:|||||||||||::||::  :||
Db        172 SPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGS 231

Qy        826 QDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHA 885
              | |||||:|||||:|||: ||||||||||||||:| ||||||||||:||  || ||||
Db        232 QYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHA 291

Qy        886 DGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGE 945
              :|||||||||||||| | :|||||||||||||||||||:|||||||| ||  :|||||
Db        292 EGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGE 351

Qy        946 RLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPAS 1004
              ||||||||||||||||||||:: ||:||||::|||:|||| |::||| :| :   |
Db        352 RLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPS 411

Qy       1005 PLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGD 1064
              | || |||:|:|::|| |:|||:|||| ||||                 | ||
Db        412 PTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF---------------NSPST------ 449

Qy       1065 LTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYS 1119
                        |:||    | |  : ::  |       :  |  |  : | :||||
Db        450 ------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYS 489

Qy       1120 EDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAK 1177
               |||      |:  ||::      | ||||||  : |: ||
Db        490 SDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPSTAM----------------- 524

Qy       1178 TLSPGKNGVVKDVFAF------------------GGAVENPEYLTPQGGAAPQPHPPPA 1218
                      |   :: |                  ||:|||||         :  |
Db        525 --------VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPL 568

Qy       1219 FSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254
                  |::  || |    :            :||:|
Db        569 AKTVFESSPYWIQSGNHQ----------INLDNPDY 594
```

```
RESULT 13
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
R;Vennstroem, B.
submitted to the EMBL Data Library, March 1993
A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X12707
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-speci
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

  Query Match             22.9%;  Score 1569;  DB 2;  Length 544;
  Best Local Similarity   52.7%;  Pred. No. 7.9e-58;
  Matches  334;  Conservative   73;  Mismatches  129;  Indels   98;  Gaps   16;

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |: |||: | | ||  ||:||  :   : :||: |     || |  |||  |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL 696
                :|||      |    || : || |:| :||:|:  |: ::||   :|| |:|||||| ||
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQEREL 113

Qy        697 VEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPV 756
              ||||||||  |||| :|||||      : : :|  :        |||||:|||:|| | |||
Db        114 VEPLTPSGEAPNQAHLRILKE------TEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPV 167

Qy        757 AIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHV 816
              ||| ||| ||||||||||||||||| | :|:| ||||||||||||:|||||||||||::
Db        168 AIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYI 227

Qy        817 RENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLL 876
              ||::  :||| |||||:|||||:|||  :|||||||||||||:| ||||||||||: |
Db        228 REHKDNIGSQYLLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAKQL 287

Qy        877 DIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPARE 936
              || ||||:||||||||||||| | :|||||||||||||||||||:|||||||| |
Db        288 GADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASE 347

Qy        937 IPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ 996
              |  :|||||||||||||||||||||||| |:: ||:||||::|||:||||| |::|||
Db        348 ISSVLEKGERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQ 407

Qy        997 -NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRS 1055
               :| :   || || |||:|:|::|| |:|||:|||| ||||                |
Db        408 GDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF---------------NS 451

Qy       1056 SSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTH 1110
               ||                    |:||      | |  : ::  | : |           |
Db        452 PST-----------------SRTPLLSSLSATSNNSATNCIDRNGG-----------H 481

Qy       1111 DPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAG 1170
                                 |    ||::      | ||||||  : |: ||
Db        482 ----------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYIS 517

Qy       1171 AT-LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203
               | : :   |  :|      :   ||:|||||
Db        518 LTAISKLPIDSRYQN-------SHSTAVDNPEYL 544


RESULT 14
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
```

```
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

  Query Match             22.8%;  Score 1562;  DB 2;  Length 545;
  Best Local Similarity   52.7%;  Pred. No. 1.5e-57;
  Matches  334;  Conservative   72;  Mismatches  130;  Indels   98;  Gaps   16;

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL 696
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQEREL 113

Qy        697 VEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPV 756
Db        114 VEPLTPSGEAPNQAHLRILKE------TEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPV 167

Qy        757 AIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHV 816
Db        168 AIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYI 227

Qy        817 RENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLL 876
Db        228 REHKDNIGSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQL 287

Qy        877 DIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPARE 936
Db        288 GADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASE 347

Qy        937 IPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ 996
Db        348 ISSVLEKGERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQ 407

Qy        997 -NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRS 1055
Db        408 GDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NS 451

Qy       1056 SSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTH 1110
Db        452 PST-----------------SRTPLLSSLSATSNNSATNCIDRNGG------------H 481

Qy       1111 DPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAG 1170
Db        482 ----------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYIS 517

Qy       1171 AT-LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203
Db        518 LTAISKLPMDSRYQN-------SHSTAVDNPEYL 544


RESULT 15
GQFFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone bi
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <LIV>
A;Cross-references: EMBL:K03054
R;Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth fac
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021
A;Molecule type: DNA
A;Residues: 'A',832-866,'V',868-943,'QTPSLVK' <WAD>
A;Cross-references: EMBL:X02293; NID:g7922; PIDN:CAA26157.1; PID:g929565
C;Comment: This sequence is tentative because the introns have not been identified
C;Genetics:
A;Gene: FlyBase:Egfr
A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2 57F
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; p
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TMM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #sta
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predi
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status p

  Query Match             22.7%;  Score 1557.5;  DB 1;  Length 1330;
  Best Local Similarity   29.0%;  Pred. No. 5.6e-57;
  Matches  404;  Conservative  178;  Mismatches  419;  Indels  391;  Gaps   42;

Qy         80 VQGYVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASP 134
Db         38 ITNYIVIGLDDLIPCTLSYRLQIIRGRTLFSLSVEEEKYALFV----------------TY 81

Qy        135 GGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHP 194
Db         82 SKMYTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPK 141

Qy        195 CSPMCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDC 252
Db        142 CHESCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDC 200

Qy        253 LACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCT 312
Db        201 IACKNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACV 259

Qy        313 LVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAG 366
Db        260 RSCPQDKMDKGGE-----CVPCNGPCPKTC--------------PGVTVLHAGNIDSFRN 300

Qy        367 CKKIFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDS 418
Db        301 CTVIDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQ 360

Qy        419 LPDLSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFV 477
Db        361 FRNLSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYV 420

Qy        478 HTVPWDQLFRNPHQALLHTANRPEDEC---------------------------------- 504
Db        421 SNIRWPAIQKEPEQKVWVNENLRADLCGKFLTILISVQHNIIMHIFAICREKWNHLLGSV 480

Qy        505 ------------------------------------------------------------ 504
Db        481 QRGRLLGSWHGSVPYLQELQFQWHLHRRLWLYIQVSINSTQDKSNEHQLTDACYSPSVPT 540
```

```
Qy       505 ------------------------------------------------------------VG 506
                                                                          :
Db       541 SLTIERARYAIQSAGLAMELEQITARSASMRHSKTLPAEGRQVPRWVFLGVCASARAGIA 600

Qy       507 EGLA-------CHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYV---N 556
             | ||        || ||      :|      |  |: : | ::|  ||       | ::
Db       601 EPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCETEC------PADHYTDEE 654

Qy       557 ARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYK------DPPF-------CVARCPSG 603
              | |   ||||    ||   | || || | :| ::|         |:       | ::||
Db       655 QRECFQRHPEC---NG---CTGPGADDCKSCRNFKLFDANETGPYVNSTMFNCTSKCPLE 708

Qy       604 VK-PDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
             ::  :  |  |  :       |   |   :    :||                :::  : | :
Db       709 MRHVNYQYTAIGPY------CAASPPRSSKITANLD------------VNMIFIITGAV 749

Qy       663 LVVVLGVVFGI-LIKRRQQKIRKYT--MRRLLQETELVEPLTPSGAMPNQAQMRILKEFN 719
             ||  : ::  :  | |::|| :| |  |   |   |  ||| ||    |  ::||:|:
Db       750 LVPTICILCVVTYICRQKQKAKKETVKMTMALSGREDSEPLRPSNIGANLCKLRIVKDAE 809

Qy       720 NFTVSFWLRVPKV-SASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAY 778
                 ||    |        ||||:|:|:||||||||||| | ::|  ::::| | |||
Db       810 -------LRKGGVLGMGAFGRVYKGVWVPEGENVKIPVAIKELLKSTGAESSEEFLREAY 862

Qy       779 VMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKG 838
             :||      : :|| :|::| : |:||||| |||||:|| || ::||: ||||  |||||
Db       863 IMASEEHVNLLKLLAVCMSSQMMLITQLMPLGCLLDYVRNNRDKIGSKALLNWSTQIAKG 922

Qy       839 MSYLEDVRLVHRDLAARNVLVK---SPNHVKITDFGLARLLDIDETEYHADGGKVPIKWM 895
             |||||: |||||||||||||:     :|    |||||:||  |  || | |||:||||:
Db       923 MSYLEEKRLVHRDLAARNVLVRLLAGEDH----DFGLAKLLSSDSNEYKAAGGKMPIKWL 978

Qy       896 ALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTI 955
             ||| |  | || :||||::|||:|||:||| :|:: |||::||||:| | :| || ||::
Db       979 ALECIRNRVFTSKSDVWAFGVTIWELLTFGQRPHENIPAKDIPDLIEVGLKLEQPEICSL 1038

Qy       956 DVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLG--PASPLDSTFYRS 1013
             |:|  :: || :|:  || |::| : |:  |||| |:: |  :     ||       | |
Db      1039 DIYCTLLSCWHLDAAMRPTFKQLTTVFAEFARDPGRYLAILGDKFTRLPA-------YTS 1091

Qy      1014 LLEDD---DMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLE 1070
               | |    :   |  | :     :  |  | |        ||:  |
Db      1092 QDEKDLIRKLAPTTDGSEAIAKPDDYLQPKAALGPS------HRTDCT------------ 1133

Qy      1071 PSEEEAP------RSPLAPSEGAGSDVFDG---DLGMGAAKGLQSLPTHDPSPLQRYSED 1121
                :| |      : |   :   | |  |    ::|:|  :
Db      1134 ---DEMPKLNRYCKDPSNKNSSTGDDERDSSAREVGVGNLR------------------- 1171

Qy      1122 PTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSP 1181
               : || : | |: | || | |    |  :      :| :  : |   |   :
Db      1172 --LDLPVDEDDYLMP-TCQPGPNNNNNMN------NPNQNNMAAVGVAAGYM-------- 1214

Qy      1182 GKNGVVKDVFAFGGAVENPEYL----TPQGGAAPQPH-----------------PPPAF 1219
                    |:     :|:|||||     |   | :| |                 | |
Db      1215 -------DLIGVPVSVDNPEYLLNAQTLGVGESPIPTQTIGIPVMGGPGTMEVKVPMPGS 1267

Qy      1220 SP-AFDNLYYWD 1230
               | : |: || |
Db      1268 EPTSSDHEYYND 1279
```

Search completed: July 22, 2003, 09:27:59
Job time : 31.3921 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.6928 Seconds
(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-695-709-14
Perfect score: 6853
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6641	96.9	1255	1	ERB2_HUMAN	P04626 homo sapien
2	5850	85.4	1257	1	ERB2_RAT	P06494 rattus norv
3	5834.5	85.1	1254	1	ERB2_MESAU	Q60553 mesocricetu
4	3063	44.7	1210	1	EGFR_HUMAN	P00533 homo sapien
5	3037	44.3	1210	1	EGFR_MOUSE	Q01279 mus musculu
6	2884.5	42.1	1308	1	ERB4_HUMAN	Q15303 homo sapien
7	2865	41.8	1308	1	ERB4_RAT	Q62956 rattus norv
8	2608.5	38.1	1167	1	XMRK_XIPMA	P13388 xiphophorus
9	2343.5	34.2	1342	1	ERB3_HUMAN	P21860 homo sapien
10	2271.5	33.1	1339	1	ERB3_RAT	Q62799 rattus norv
11	1879	27.4	1426	1	EGFR_DROME	P04412 drosophila
12	1668.5	24.3	634	1	ERBB_ALV	P00534 avian leuko
13	1622	23.7	604	1	ERBB_AVIER	P00535 avian eryth
14	1595	23.3	703	1	EGFR_CHICK	P13387 gallus gall
15	1552	22.6	540	1	ERBB_AVIEU	P11273 avian eryth
16	1241	18.1	1323	1	LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.7	245	1	ERB2_MOUSE	P70424 mus musculu
18	700	10.2	1363	1	ILPR_BRALA	O02466 branchiosto
19	670	9.8	1383	1	INSR_RAT	P15127 rattus norv
20	669.5	9.8	1372	1	INSR_MOUSE	P15208 mus musculu
21	665	9.7	1300	1	IRR_MOUSE	Q9wtl4 mus musculu
22	665	9.7	1382	1	INSR_HUMAN	P06213 homo sapien
23	656	9.6	1297	1	IRR_HUMAN	P14616 homo sapien
24	655	9.6	1477	1	HTK7_HYDAT	Q25197 hydra atten
25	653	9.5	1607	1	MIPR_LYMST	Q25410 lymnaea sta
26	651.5	9.5	1300	1	IRR_CAVPO	P14617 cavia porce
27	596	8.7	1367	1	IG1R_HUMAN	P08069 homo sapien
28	590	8.6	1390	1	INSR_AEDAE	Q93105 aedes aegyp
29	582	8.5	2146	1	INSR_DROME	P09208 drosophila
30	578	8.4	1373	1	IG1R_MOUSE	Q60751 mus musculu
31	574.5	8.4	1370	1	IG1R_RAT	P24062 rattus norv
32	574	8.4	1114	1	RET_HUMAN	P07949 homo sapien
33	571	8.3	987	1	EPB4_HUMAN	P54760 homo sapien
34	569.5	8.3	984	1	EPB1_CHICK	Q07494 gallus gall
35	566.5	8.3	984	1	EPB1_RAT	P09759 rattus norv
36	562.5	8.2	977	1	EPA2_MOUSE	Q03145 mus musculu
37	560.5	8.2	984	1	EPB1_HUMAN	P54762 homo sapien
38	558	8.1	902	1	EPBB_XENLA	Q91736 xenopus lae
39	556.5	8.1	976	1	EPA2_HUMAN	P29317 homo sapien
40	551	8.0	987	1	EPB4_MOUSE	P54761 mus musculu
41	550.5	8.0	985	1	EPBA_XENLA	Q91571 xenopus lae
42	544	7.9	1068	1	FAK1_XENLA	Q91738 xenopus lae
43	539.5	7.9	1053	1	FAK1_CHICK	Q00944 gallus gall
44	537	7.8	757	1	HT16_HYDAT	P53356 hydra atten
45	532.5	7.8	1052	1	FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1
ERB2_HUMAN
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

```
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
CC       POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC       OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC       ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11767; AAA35808.1; -.
DR   EMBL; M11761; AAA35808.1; JOINED.
DR   EMBL; M11762; AAA35808.1; JOINED.
DR   EMBL; M11763; AAA35808.1; JOINED.
DR   EMBL; M11764; AAA35808.1; JOINED.
DR   EMBL; M11765; AAA35808.1; JOINED.
DR   EMBL; M11766; AAA35808.1; JOINED.
DR   EMBL; M11730; AAA75493.1; -.
DR   EMBL; M12036; AAA35978.1; -.
DR   EMBL; X03363; CAA27060.1; -.
DR   PIR; A25491; A25491.
DR   PIR; A24571; A24571.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3430; ERBB2.
DR   MIM; 164870; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Polymorphism.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   1255       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN      22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   653    675       POTENTIAL.
FT   DOMAIN     676   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     720    987       PROTEIN KINASE.
FT   NP_BIND    726    734       ATP (BY SIMILARITY).
FT   BINDING    753    753       ATP (BY SIMILARITY).
FT   ACT_SITE   845    845       BY SIMILARITY.
FT   DISULFID   195    204       BY SIMILARITY.
FT   DISULFID   199    212       BY SIMILARITY.
FT   DISULFID   220    227       BY SIMILARITY.
FT   DISULFID   224    235       BY SIMILARITY.
FT   DISULFID   236    244       BY SIMILARITY.
FT   DISULFID   240    252       BY SIMILARITY.
FT   DISULFID   255    264       BY SIMILARITY.
FT   DISULFID   268    295       BY SIMILARITY.
FT   DISULFID   299    311       BY SIMILARITY.
FT   DISULFID   315    331       BY SIMILARITY.
FT   DISULFID   334    338       BY SIMILARITY.
FT   DISULFID   511    520       BY SIMILARITY.
FT   DISULFID   515    528       BY SIMILARITY.
FT   DISULFID   531    540       BY SIMILARITY.
FT   DISULFID   544    560       BY SIMILARITY.
FT   DISULFID   563    576       BY SIMILARITY.
FT   DISULFID   567    584       BY SIMILARITY.
FT   DISULFID   587    596       BY SIMILARITY.
FT   DISULFID   600    623       BY SIMILARITY.
FT   DISULFID   626    634       BY SIMILARITY.
FT   DISULFID   630    642       BY SIMILARITY.
FT   MOD_RES   1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    654    654       I -> V.
FT                               /FTId=VAR_004077.
FT   VARIANT    655    655       I -> V.
FT                               /FTId=VAR_004078.
FT   CONFLICT  1170   1170       P -> A (IN REF. 2).
SQ   SEQUENCE  1255 AA;  137909 MW;  39E9DFDA04DCF962 CRC64;

  Query Match             96.9%;  Score 6641;  DB 1;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719
```

```
Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||  |    ||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 2
ERB2_RAT
ID   ERB2_RAT       STANDARD;      PRT;  1257 AA.
AC   P06494;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE   receptor-related protein).
GN   ERBB2 OR NEU.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=86118662; PubMed=3945311;
RA   Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT   "The neu oncogene encodes an epidermal growth factor receptor-related
RT   protein.";
RL   Nature 319:226-230(1986).
RN   [2]
RP   SEQUENCE OF 852-905 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   STRUCTURE BY NMR OF 650-668.
RX   MEDLINE=92155181; PubMed=1346763;
RA   Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,
RA   Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT   "Three dimensional structure of the transmembrane region of the proto-
RT   oncogenic and oncogenic forms of the neu protein.";
RL   EMBO J. 11:43-48(1992).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC       THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03362; CAA27059.1; ALT_INIT.
DR   PIR; A24562; TVRTNU.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   1257       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN      22    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   655    677       POTENTIAL.
FT   DOMAIN     678   1257       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     159    369       CYS-RICH.
FT   DOMAIN     473    646       CYS-RICH.
FT   DOMAIN     722    989       PROTEIN KINASE.
FT   NP_BIND    728    736       ATP (BY SIMILARITY).
FT   BINDING    755    755       ATP (BY SIMILARITY).
FT   ACT_SITE   847    847       BY SIMILARITY.
FT   DISULFID   196    205       BY SIMILARITY.
FT   DISULFID   200    213       BY SIMILARITY.
FT   DISULFID   221    228       BY SIMILARITY.
FT   DISULFID   225    236       BY SIMILARITY.
FT   DISULFID   237    245       BY SIMILARITY.
FT   DISULFID   241    253       BY SIMILARITY.
FT   DISULFID   256    265       BY SIMILARITY.
FT   DISULFID   269    296       BY SIMILARITY.
FT   DISULFID   300    312       BY SIMILARITY.
FT   DISULFID   316    332       BY SIMILARITY.
FT   DISULFID   335    339       BY SIMILARITY.
FT   DISULFID   513    522       BY SIMILARITY.
```

```
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    578       BY SIMILARITY.
FT   DISULFID    569    586       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    602    625       BY SIMILARITY.
FT   DISULFID    628    636       BY SIMILARITY.
FT   DISULFID    632    644       BY SIMILARITY.
FT   MOD_RES    1141   1141       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1250   1250       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    260    260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     661    661       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1257 AA;  138831 MW;  6129264583011402 CRC64;

  Query Match             85.4%;  Score 5850;  DB 1;  Length 1257;
  Best Local Similarity   85.8%;  Pred. No. 3.2e-306;
  Matches 1084;  Conservative   52;  Mismatches  118;  Indels   10;  Gaps    4;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||| |||||||| | |||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| |||||||||||||||:||||||::||||||||||||||||| ||||||||
Db         61 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||: |||||||| :||||:| |||||
Db        121 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 180

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |  |||||||| ||:| || : ||||| |||| || |:|  ||||||| |||||||||
Db        181 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 240

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||| |
Db        241 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 300

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||||||||:::::
Db        301 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 360

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
                || |||||||||||||||||||:|  |||:|||||||||||||||||||||||||||
Db        361 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 420

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              ||||||||::|||||||:|||||||||||  |||||||||||||||| | |||||||
Db        421 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 480

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||: ||||:: ||  || |: ||| ||||||||||||||| |||||
Db        481 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : |||||||||||| | |||| |||| |||||||   |||
Db        541 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 600

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAV 658
              |||||||||||||||||:|||| ||||||||||||||::||||||||:| |::  |
Db        601 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRASPVTFIIATV 660

Qy        659 VGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEF 718
              ||:|| ::|  || |||||||:|||||||||||||||||||||||||||||||||||||
Db        661 VGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 720

Qy        719 WNFTVSPWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 777
                                ||  ||  |  ||||||||||||||||||||||||||||||||||||
Db        721 E-------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 773

Qy        778 YVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAK 837
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||:||||
Db        774 YVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAK 833

Qy        838 GMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMAL 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 GMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMAL 893

Qy        898 ESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDV 957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDV 953

Qy        958 YMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLED 1017
              ||||||||||||||||||||||||||||||||||||||||||||||:||:||||||||||
Db        954 YMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLED 1013

Qy       1018 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAP 1077
              |||||||||||||||||| ||| || |   | ||||||||||:|||||||||| |
Db       1014 DDMGDLVDAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPP 1073

Qy       1078 RSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPL 1137
              |||||||||||||||||| || |||||| || ||||||||||:||| |||||||||
Db       1074 RSPLAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPL 1133

Qy       1138 TCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAV 1197
              |||||||||| :|:|||| |||||  |||||||| ||||||||||||||||||||||||
Db       1134 ACSPQPEYVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAV 1193

Qy       1198 ENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGL 1257
              |||||| |: | |  ||| |||||||||||||||:  |:| ||| |:|||||||||||||
Db       1194 ENPEYLVPREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGL 1253

Qy       1258 DVPV 1261
              ||||
Db       1254 DVPV 1257
```

```
RESULT 3
ERB2_MESAU
ID   ERB2_MESAU     STANDARD;      PRT;  1254 AA.
AC   Q60553;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN   ERBB2 OR NEU.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Nerve;
RX   MEDLINE=94193007; PubMed=7908275;
RA   Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA   Yamazaki Y., Ishikawa T.;
RT   "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL   Gene 140:251-255(1994).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
```

```
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D16295; BAA03801.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1254       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1254       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      158    368       CYS-RICH.
FT   DOMAIN      472    644       CYS-RICH.
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1247   1247       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     658    658       V -> E (IN ONCOGENIC NEU).
FT   VARIANT     659    659       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1254 AA;  138252 MW;  974C3791C21F2BE1 CRC64;
```

```
  Query Match             85.1%;  Score 5834.5;  DB 1;  Length 1254;
  Best Local Similarity   85.3%;  Pred. No. 2.2e-305;
  Matches 1077;  Conservative   61;  Mismatches  115;  Indels    9;  Gaps    3;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| | |||||||| ||| |||||||||||||||||||||||::||||||||||||||
Db          1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:||||:||| |||||||||||||||| ||||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||:| |  || :| |||||||||||||||||||: |||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                |||||||| ||:| || : ||| | ||||:|| |:         :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||:|||||||||:|||||||||||||||||| ||||||||||::::
Db        301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||:|:|  ||| |||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              |||||||||:|||||:||:||||| |||||| |||||||||||||| ||| ||||||||||
Db        421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||: |  |:||  :  ||: ||| |||||||||||||| |||||||
Db        481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |:|||| :||||||||:|||||||||||| : || | |||| || |||| |||||||
Db        541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              |||||||||||||||:|||| ||||||||||||||||::|||||||||| |||:: |||
Db        601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||| :|:||| |||||||:|||||||||||||||||||||||||||||||||||||||||
Db        661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                    ||  ||  |     ||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              |||:|||||||||||||||||||||||||||||||||||:|||||||||||||:||||||
Db        774 MAGLGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDD 1013
```

```
Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||| ||||||||    | ||||||||||||:||||:||| || |||
Db       1014 MGDLVDAEEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||:|:||||| || ||:   | ||||||||||:|||:||||||||| |
Db       1074 PLAPSEGAGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLAC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              ||||||||||:||||||   |||||  |||||||| |||||||||||||||| |||||||
Db       1134 SPQPEYVNQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              |||| |:||:| ||| |||  |||||||||||||| |||:||:||:|||||||||||||||
Db       1194 PEYLVPRGGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDV 1252

Qy       1260 PV 1261
              ||
Db       1253 PV 1254
```

RESULT 4
EGFR_HUMAN
ID   EGFR_HUMAN   STANDARD;   PRT; 1210 AA.
AC   P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC   O00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE   protein-tyrosine kinase ErbB-1).
GN   EGFR OR ERBB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=84219729; PubMed=6328312;
RA   Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA   Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA   Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT   "Human epidermal growth factor receptor cDNA sequence and aberrant
RT   expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL   Nature 309:418-425(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=95382957; PubMed=7654368;
RA   Ilekis J.V., Stark B.C., Scoccia B.;
RT   "Possible role of variant RNA transcripts in the regulation of
RT   epidermal growth factor receptor expression in human placenta.";
RL   Mol. Reprod. Dev. 41:149-156(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97078686; PubMed=8918811;
RA   Reiter J.L., Maihle N.J.;
RT   "A 1.8 kb alternative transcript from the human epidermal growth
RT   factor receptor gene encodes a truncated form of the receptor.";
RL   Nucleic Acids Res. 24:4050-4056(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97256547; PubMed=9103388;
RA   Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT   "Expression of a truncated epidermal growth factor receptor-like
RT   protein (TEGFR) in ovarian cancer.";
RL   Gynecol. Oncol. 65:36-41(1997).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC   TISSUE=Placenta;
RX   MEDLINE=21100872; PubMed=11161793;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative EGFR transcripts encoding truncated receptor
RT   isoforms.";
RL   Genomics 71:1-20(2001).
RN   [6]
RP   SEQUENCE OF 575-687 FROM N.A.
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Human and mouse alternative EGFR transcripts encoding only the
RT   extracellular domain of the receptor.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 713-924 FROM N.A.
RX   MEDLINE=84196372; PubMed=6326261;
RA   Lin C.R., Chen W.S., Kruiger W., Stolarsky L.S., Weber W.,
RA   Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT   "Expression cloning of human EGF receptor complementary DNA: gene
RT   amplification and three related messenger RNA products in A431
RT   cells.";
RL   Science 224:843-848(1984).
RN   [8]
RP   SEQUENCE OF 150-962 FROM N.A.
RX   MEDLINE=84245835; PubMed=6330563;
RA   Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA   Roe B.A., Merlino G.T., Pastan I.;
RT   "Human epidermal growth factor receptor cDNA is homologous to a
RT   variety of RNAs overproduced in A431 carcinoma cells.";
RL   Nature 309:806-810(1984).
RN   [9]
RP   SEQUENCE OF 1028-1210 FROM N.A.
RX   MEDLINE=85046483; PubMed=6093780;
RA   Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA   O'Malley B.W.;
RT   "Isolation of an evolutionarily conserved epidermal growth factor
RT   receptor cDNA from human A431 carcinoma cells.";
RL   Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN   [10]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=88217333; PubMed=3329716;
RA   Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA   Waterfield M.D.;
RT   "The human EGF receptor gene: structure of the 110 kb locus and
RT   identification of sequences regulating its transcription.";
RL   Oncogene Res. 1:375-396(1987).
RN   [11]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=91107677; PubMed=1988448;
RA   Haley J.D., Waterfield M.D.;
RT   "Contributory effects of de novo transcription and premature
RT   transcript termination in the regulation of human epidermal growth
RT   factor receptor proto-oncogene RNA synthesis.";
RL   J. Biol. Chem. 266:1746-1753(1991).
RN   [12]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=85270438; PubMed=2991899;
RA   Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT   "Characterization and sequence of the promoter region of the human
RT   epidermal growth factor receptor gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN   [13]
RP   SEQUENCE OF 540.
RA   Kohda D.;
RL   Submitted (SEP-1997) to the SWISS-PROT data bank.
RN   [14]
RP   RECEPTOR ACTIVITY.
RX   MEDLINE=84191554; PubMed=6325948;
RA   Mroczkowski B., Mosig G., Cohen S.;
RT   "ATP-stimulated interaction between epidermal growth factor receptor

```
RT   and supercoiled DNA.";
RL   Nature 309:270-273(1984).
RN   [15]
RP   PHOSPHORYLATION.
RX   MEDLINE=89278137; PubMed=2543678;
RA   Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA   Howk R., Givol D., Ullrich A., Schlessinger J.;
RT   "All autophosphorylation sites of epidermal growth factor (EGF)
RT   receptor and HER2/neu are located in their carboxyl-terminal tails.
RT   Identification of a novel site in EGF receptor.";
RL   J. Biol. Chem. 264:10667-10671(1989).
RN   [16]
RP   CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
RP   ASN-528.
RX   MEDLINE=96398132; PubMed=8962717;
RA   Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
RT   "Analysis of the glycosylation patterns of the extracellular domain of
RT   the epidermal growth factor receptor expressed in Chinese hamster
RT   ovary fibroblasts.";
RL   Growth Factors 13:121-132(1996).
RN   [17]
RP   CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
RP   ASN-603.
RX   MEDLINE=20198209; PubMed=10731668;
RA   Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
RT   "Characterization of the N-oligosaccharides attached to the atypical
RT   Asn-X-Cys sequence of recombinant human epidermal growth factor
RT   receptor.";
RL   J. Biochem. 127:65-72(2000).
RN   [18]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=98225196; PubMed=9556602;
RA   Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
RT   "Disulfide bond structure of human epidermal growth factor receptor.";
RL   J. Biol. Chem. 273:11150-11157(1998).
RN   [19]
RP   REVIEW.
RX   MEDLINE=87297456; PubMed=3039909;
RA   Carpenter G.;
RT   "Receptors for epidermal growth factor and other polypeptide
RT   mitogens.";
RL   Annu. Rev. Biochem. 56:881-914(1987).
CC   -!- FUNCTION: Receptor for EGF, but also for other members of the EGF
CC       family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
CC       EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
CC       involved in the control of cell growth and differentiation.
CC   -!- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
CC       secreted.
CC   -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
CC       truncated isoform/TEGFR, 3/p110 and 4; are produced by
CC       alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
CC       expressed in ovarian cancers.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00588; CAA25240.1; -.
DR   EMBL; U95089; AAB53063.1; -.
DR   EMBL; U48722; AAC50802.1; -.
DR   EMBL; U48723; AAC50804.1; -.
DR   EMBL; U48724; AAC50796.1; -.
DR   EMBL; U48725; AAC50797.1; -.
DR   EMBL; U48726; AAC50798.1; -.

  Query Match             44.7%;  Score 3063;  DB 1;  Length 1210;
  Best Local Similarity   48.7%;  Pred. No. 7e-157;
  Matches  620;  Conservative  178;  Mismatches  360;  Indels  114;  Gaps   24;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
Db        127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy        189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
Db        184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy        248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
Db        244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy        308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
Db        304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy        367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
Db        362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy        427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
Db        422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy        487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
Db        482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNL 541

Qy        547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606
Db        542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy        607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG---ILL 663
Db        602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLL 658

Qy        664 VVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNNFTV 723
Db        659 VVALGIG---LFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKE------ 709

Qy        724 SFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGV 783
Db        710 TEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASV 769

Qy        784 GSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLE 843
Db        770 DNPHVCRLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLE 829

Qy        844 DVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRR 903
Db        830 DRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHR 889

Qy        904 RFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVK 963
Db        890 IYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVK 949
```

```
Qy        964 CWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGD 1022
              |||||:: ||:|||| ||||:|||||||::||| :| :   || || |||:|::::|| |
Db        950 CWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDD 1009

Qy       1023 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1082
              :|||:|||:|||||                                    |     |:||
Db       1010 VVDADEYLIPQQGFF---------------------------------SSPSTSRTPLL 1035

Qy       1083 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCS 1140
               |  | |:  :  :      ||||  |  : | ||||| |||  |  ::  | ::
Db       1036 SSLSATSN--NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------ 1087

Qy       1141 PQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENP 1200
              | |||:||  |  :|    : |:   :|           :| ::   :|      || ||
Db       1088 PVPEYINQ-SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNP 1136

Qy       1201 EYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTF 1243
              ||| | |           |   :  ||:  :| |         :|       |:   |   |
Db       1137 EYLNTVQ---------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIF 1187

Qy       1244 KGTPTAENPEYL 1255
              ||: |||| |||
Db       1188 KGS-TAENAEYL 1198
```

```
RESULT 5
EGFR_MOUSE
ID   EGFR_MOUSE     STANDARD;      PRT;  1210 AA.
AC   Q01279;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=93026370; PubMed=1408137;
RA   Avivi A., Skorecki K., Yayon A., Givol D.;
RT   "Promoter region of the murine fibroblast growth factor receptor 2
RT   (bek/KGFR) gene.";
RL   Oncogene 7:1957-1962(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX   MEDLINE=93126380; PubMed=7678348;
RA   Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT   "Expression of the epidermal growth factor receptor gene is regulated
RT   in mouse blastocysts during delayed implantation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RA   Hibbs M.L.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/C3; TISSUE=Liver;
RX   MEDLINE=94170986; PubMed=8125255;
RA   Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA   Jenkins N.A., Lee D.C.;
RT   "The mouse waved-2 phenotype results from a point mutation in the EGF
RT   receptor tyrosine kinase.";
RL   Genes Dev. 8:399-413(1994).
RN   [5] .
RP   SEQUENCE OF 1-714 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=91232866; PubMed=2030916;
RA   Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT   "Comparison of EGF receptor sequences as a guide to study the ligand
RT   binding site.";
RL   Oncogene 6:673-676(1991).
RN   [6]
RP   SEQUENCE OF 969-1117 FROM N.A.
RC   STRAIN=C3H;
RA   Eisinger D.P., Serrero G.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78987; CAA55587.1; -.
DR   EMBL; U03425; AAA17899.1; -.
DR   EMBL; X59698; CAA42219.1; -.
DR   EMBL; L06864; AAA53029.1; -.
DR   EMBL; Z12608; CAA78249.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1210       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       25    647       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    648    670       POTENTIAL.
FT   DOMAIN      671   1210       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       75    300       APPROXIMATE.
FT   REPEAT      390    600       APPROXIMATE.
FT   DOMAIN     1028   1071       SER-RICH.
FT   DOMAIN      714    981       PROTEIN KINASE.
FT   NP_BIND     720    728       ATP (BY SIMILARITY).
FT   BINDING     747    747       ATP (BY SIMILARITY).
FT   ACT_SITE    839    839       BY SIMILARITY.
FT   DISULFID    190    199       BY SIMILARITY.
FT   DISULFID    194    207       BY SIMILARITY.
FT   DISULFID    215    223       BY SIMILARITY.
FT   DISULFID    219    231       BY SIMILARITY.
FT   DISULFID    232    240       BY SIMILARITY.
FT   DISULFID    236    248       BY SIMILARITY.
FT   DISULFID    251    260       BY SIMILARITY.
FT   DISULFID    264    291       BY SIMILARITY.
FT   DISULFID    295    307       BY SIMILARITY.
FT   DISULFID    311    326       BY SIMILARITY.
FT   DISULFID    329    333       BY SIMILARITY.
```

```
FT   DISULFID    506    515       BY SIMILARITY.
FT   DISULFID    510    523       BY SIMILARITY.
FT   DISULFID    526    535       BY SIMILARITY.
FT   DISULFID    539    555       BY SIMILARITY.
FT   DISULFID    558    571       BY SIMILARITY.
FT   DISULFID    562    579       BY SIMILARITY.
FT   DISULFID    582    591       BY SIMILARITY.
FT   DISULFID    595    617       BY SIMILARITY.
FT   DISULFID    620    628       BY SIMILARITY.
FT   DISULFID    624    636       BY SIMILARITY.
FT   MOD_RES     680    680       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   MOD_RES    1092   1092       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1110   1110       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1172   1172       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1197   1197       PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT                                (BY SIMILARITY).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    175    175       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    352    352       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    444    444       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     19     19       C -> S (IN REF. 2).
FT   CONFLICT    539    539       C -> W (IN REF. 5).
FT   CONFLICT    991    991       L -> F (IN REF. 4).
FT   CONFLICT   1116   1117       HP -> DR (IN REF. 6).
SQ   SEQUENCE   1210 AA;  134853 MW;  690E20D46DF2D2F5 CRC64;

  Query Match             44.3%;  Score 3037;  DB 1;  Length 1210;
  Best Local Similarity   48.6%;  Pred. No. 1.7e-155;
  Matches  622;  Conservative  172;  Mismatches  367;  Indels  118;  Gaps   26;

Qy          11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
               ||  ||    | |     :|| ||  :|    : | |    |: :|  |:|| ||||:||:  |
Db          14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy          69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
               ||||: |||| |||||| | | ::||: |:|:|| |:|: ||||:| |
Db          74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy         129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
                 | :  ||||| :|:| ||| | |    || ||  ||| |:||   |  : |
Db         125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy         185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
               :  :|  | | |    |||  |:|| ||: :||  |: ||:|  |:|||| |||||
Db         181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy         244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
               ||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:||  || ||
Db         240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy         304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
               : || |||    |      ||  ||| ::|:||  || :|| |:|:   | ::  |  | ::
Db         300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy         364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
                |  | | |  | || :| ||  :  | || | :|:: :|::|||:| | ||||:  ||
Db         358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy         423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
               |:||::||||   :| :|| : || |: ||||||:|:  |  :|  | :||: :|: |
Db         418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy         483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
               :||  |:|      || | :|     |: ||:   |||| |  ||:|    ||:||||
Db         478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy         543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
```

```
               :| :|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||  ||:
Db         538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy         603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
               |:  : : : :||: |    |  | |||: |     :||       : || : :|| |
Db         598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy         663 LVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNNF 721
               | :|: |  || |  ||:  :|| |:|||||| |||||||||  |||| :|||||
Db         657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKE---- 711

Qy         722 TVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 781
                 : : ::  : :    |||||:|||:|| |||||||| ||| ||||||||||||||||
Db         712 --TEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMA 769

Qy         782 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 841
               | :|:| |||||||||||||:|||||||||||:|||::  :||| |||||:|||||:|
Db         770 SVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNY 829

Qy         842 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 901
               ||| |||||||||||||||:| ||||||||||:||  :| |||||:|||||||||||||
Db         830 LEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESIL 889

Qy         902 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 961
               | :|||||||||||||||||||:|||||||| :|  :|||||||||||||||||||||
Db         890 HRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIM 949

Qy         962 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDM 1020
               |||||||:: ||:||||: |||:|||||||::||| :| :   || || |||:|::::||
Db         950 VKCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDM 1009

Qy        1021 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1080
                |:|||:|||:|||||                  | ||                   |:|
Db        1010 EDVVDADEYLIPQQGFF----------------NSPST-----------------SRTP 1035

Qy        1081 LAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLT 1138
               |  |  | |:    :  :       |    : : |||||  |||  :  :   | ::
Db        1036 LLSSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL---- 1087

Qy        1139 CSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVE 1198
                  | ||||||  |  :|    : |:  :|           :||::  ::       ||
Db        1088 --PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVG 1134

Qy        1199 NPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPS 1241
               ||||| | |          |  |  |::   | |          :|        |:   |
Db        1135 NPEYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNG 1185

Qy        1242 TFKGTPTAENPEYLGLDVP 1260
                ||| ||||| ||| :  |
Db        1186 IFKG-PTAENAEYLRVAPP 1203


RESULT 6
ERB4_HUMAN
ID   ERB4_HUMAN     STANDARD;      PRT;  1308 AA.
AC   Q15303;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE   (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN   ERBB4 OR HER4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM JM-A).
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=93189574; PubMed=8383326;
RA   Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA   Foy L., Neubauer M.G., Shoyab M.;
RT   "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
```

```
RT   epidermal growth factor receptor family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97476287; PubMed=9334263;
RA   Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA   Klagsbrun M.;
RT   "A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
RT   tissue distribution and differential processing in response to
RT   phorbol ester.";
RL   J. Biol. Chem. 272:26761-26768(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC       FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC       BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC       NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC       CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC       KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC       PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC       LUNG, SALIVARY GLAND, AND PANCREAS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07868; AAB59446.1; -.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3432; ERBB4.
DR   MIM; 600543; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN      676   1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      186    334       CYS-RICH.
FT   DOMAIN      496    633       CYS-RICH.
FT   DOMAIN      718    985       PROTEIN KINASE.
FT   NP_BIND     724    732       ATP (BY SIMILARITY).
FT   BINDING     751    751       ATP (BY SIMILARITY).
FT   ACT_SITE    843    843       BY SIMILARITY.
FT   DISULFID    189    197       BY SIMILARITY.
FT   DISULFID    193    205       BY SIMILARITY.
FT   DISULFID    213    221       BY SIMILARITY.
FT   DISULFID    217    229       BY SIMILARITY.
FT   DISULFID    230    238       BY SIMILARITY.
FT   DISULFID    234    246       BY SIMILARITY.
FT   DISULFID    249    258       BY SIMILARITY.
FT   DISULFID    262    289       BY SIMILARITY.
FT   DISULFID    293    304       BY SIMILARITY.
FT   DISULFID    308    323       BY SIMILARITY.
FT   DISULFID    326    330       BY SIMILARITY.
FT   DISULFID    503    512       BY SIMILARITY.
FT   DISULFID    507    520       BY SIMILARITY.
FT   DISULFID    523    532       BY SIMILARITY.
FT   DISULFID    536    552       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    559    577       BY SIMILARITY.
FT   DISULFID    580    589       BY SIMILARITY.
FT   DISULFID    593    614       BY SIMILARITY.
FT   DISULFID    617    625       BY SIMILARITY.
FT   DISULFID    621    633       BY SIMILARITY.
FT   MOD_RES    1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    626    648       NGPTSHDCIYYPWTGHSTLPQHA -> IGSSIEDCIGLMD
FT                                (IN ISOFORM JM-B).
SQ   SEQUENCE   1308 AA;  146807 MW;  5E4AE80985D88761 CRC64;

  Query Match           42.1%;  Score 2884.5;  DB 1;  Length 1308;
  Best Local Similarity   44.2%;  Pred. No. 2.8e-147;
  Matches  598;  Conservative  185;  Mismatches  390;  Indels  179;  Gaps   29;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  | ||||: ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::||||| ||| : :|  ||| ||| |:||       |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |  | | |||| :  ||:|||||| |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||    ||  ||:  |   :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||    ||  |:| : |: |:  | : | |:|   : :          :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :   : ||:| || |: ||||:| | :|| :: | |
```

```
Db       356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy       424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483

Db       416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy       484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543

Db       476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy       544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602

Db       536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy       603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PAEQRASPL 651

Db       596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQHAR-TPL 652

Qy       652 TSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Db       653 --IAAGVIGGLFILVIVGLTFAVYVRRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNQA 708

Qy       711 QMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKAN 770

Db       709 QLRILKE------TELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKAN 762

Qy       771 KEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLN 830

Db       763 VEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLLN 822

Qy       831 WCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKV 890

Db       823 WCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKM 882

Qy       891 PIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQP 950

Db       883 PIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQP 942

Qy       951 PICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDST 1009

Db       943 PICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDSK 1002

Qy      1010 FYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGL 1069

Db      1003 FFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIGH 1050

Qy      1070 EPSEEEAPRS----------------------------PLAP-SEGAGSDVFDGDLGMGA 1100

Db      1051 SPPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNGT 1110

Qy      1101 AKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRP 1153

Db      1111 LRKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE--- 1167

Qy      1154 QPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQP 1213

Db      1168 ------ENPFVSRR---------------KNGDLQ-------ALDNPEYHNASNG----- 1194

Qy      1214 HPPPA----------------------------------FSPAFDNLYYWDQDPPERGA- 1238

Db      1195 -PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRSTL 1253

Qy      1239 -PPSTFKGTPT--------------AENPEYL 1255

Db      1254 QHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 7
ERB4_RAT
ID   ERB4_RAT       STANDARD;      PRT;  1308 AA.
AC   Q62956; Q9Z2N7;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN   ERBB4 OR TYRO-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=98221155; PubMed=9553078;
RA   Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA   Marchionni M.A., Kelly R.A.;
RT   "Neuregulins promote survival and growth of cardiac myocytes.
RT   Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT   ventricular myocytes.";
RL   J. Biol. Chem. 273:10261-10269(1998).
RN   [2]
RP   SEQUENCE OF 848-901 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   SEQUENCE OF 1031-1198 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC       NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC       OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC       RETICULAR NUCLEUS OF THE THALLAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC       HEART.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF041838; AAD08899.1; -.
DR   EMBL; U52531; AAC53051.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
```

```
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN      676   1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      186    334       CYS-RICH.
FT   DOMAIN      496    633       CYS-RICH.
FT   DOMAIN      718    985       PROTEIN KINASE.
FT   NP_BIND     724    732       ATP (BY SIMILARITY).
FT   BINDING     751    751       ATP (BY SIMILARITY).
FT   ACT_SITE    843    843       BY SIMILARITY.
FT   DISULFID    189    197       BY SIMILARITY.
FT   DISULFID    193    205       BY SIMILARITY.
FT   DISULFID    213    221       BY SIMILARITY.
FT   DISULFID    217    229       BY SIMILARITY.
FT   DISULFID    230    238       BY SIMILARITY.
FT   DISULFID    234    246       BY SIMILARITY.
FT   DISULFID    249    258       BY SIMILARITY.
FT   DISULFID    262    289       BY SIMILARITY.
FT   DISULFID    293    304       BY SIMILARITY.
FT   DISULFID    308    323       BY SIMILARITY.
FT   DISULFID    326    330       BY SIMILARITY.
FT   DISULFID    503    512       BY SIMILARITY.
FT   DISULFID    507    520       BY SIMILARITY.
FT   DISULFID    523    532       BY SIMILARITY.
FT   DISULFID    536    552       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    559    577       BY SIMILARITY.
FT   DISULFID    580    589       BY SIMILARITY.
FT   DISULFID    593    614       BY SIMILARITY.
FT   DISULFID    617    625       BY SIMILARITY.
FT   DISULFID    621    633       BY SIMILARITY.
FT   MOD_RES    1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1062   1062       S -> N (IN REF. 3).
FT   CONFLICT   1080   1082       PYT -> SYR (IN REF. 3).
SQ   SEQUENCE   1308 AA;  146957 MW;  D944BB0996A08B41 CRC64;

  Query Match             41.8%;  Score 2865;  DB 1;  Length 1308;
  Best Local Similarity   44.1%;  Pred. No. 3.1e-146;
  Matches  596;  Conservative  192;  Mismatches  399;  Indels  166;  Gaps   29;

Qy          1 MELA-ALCRWGLLL--ALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQ 57
              |:||  |  || ||  |     :||  || ||: ||   :  |     ||  |: |:||
Db          1 MKLATGLWVWGSLLVAARTVQPSASQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVM 60

Qy         58 GNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVL 117
              ||||:| :   |  ||||: |:|| ||||:| || | :||: |||:|||:|:|| ||||:
Db         61 GNLEITSIEHNRDLSFLRSIREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIF 120

Qy        118 DNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNN 177
              |     |          ||:|| |::||||| ||| : :|  ||| ||| |:||
Db        121 LNYRKDGNF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPW 171

Qy        178 QLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCC 236
               :||: |  |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||
Db        172 PSNMTLVSTIGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCC 230

Qy        237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
              | :|| ||:||| :|| ||::|| || |   ||     ||  ||:    |  :||:|| ||
Db        231 HRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCV 290

Qy        297 TACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF 356
               ||:|::  |  ||   ||    ||  |:| : |: |:  | : | |:|   : :
Db        291 KKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTV 348

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
                 :|  | || |:| ||      ||| :    : ||:| || |: |||:| |  ||
Db        349 DSSNIDKFINCTKINGNLIFLVTGIHGDPYNAIDAIDPEKLNVFRTVREITGFLNIQTWP 408

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
              :: | ||| ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||:
Db        409 PNMTDFSVFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCY 468

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
               ||: |  ||   :| ::   ||  : |  ||: |: ||:   |||||| ||::| :| |
Db        469 YHTINWTTLFSTVNQRIVIRDNRRAENCTAEGMVCNHLCSNDGCWGPGPDQCLSCRRFSR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPF 595
              |: |:|  | :  |  ||: |   |: |  :|:  ::| :|| ||  | |  |:|:|| |
Db        529 GKICIESCNLYDGEFREFENGSICVECDSQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPN 588

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PA 644
              || :||  ::   |:  |:|: |::  | ||  |||  |        |          |
Db        589 CVEKCPDVLQGANSF--IFKYADQDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQ 646

Qy        645 EQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPS 703
               | :||  | : |: |: ::|:: : | : ::|:  | :|  :|| | ||||||||||
Db        647 HAR-TPL--IAAGVIGGLFILVIMALTFAVYVRRKSIK-KKRALRRFL-ETELVEPLTPS 701

Qy        704 GAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRE 763
              |  |||||:|||||      :   ||  : :    |||||||:|:|| ||||||||:| |
Db        702 GTAPNQAQLRILKE------TELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNE 755

Qy        764 NTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRL 823
               | |||| | :||| :|| |   |:: ||||:||: |:|||||||:||||::| |::  :
Db        756 TTGPKANVEFMDEALIMASVDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNI 815

Qy        824 GSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEY 883
              ||| |||||:|||||| |||: ||||||||||||||||||||||||||||||: || ||
Db        816 GSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEY 875

Qy        884 HADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 943
              :|||||:|||||||| |  |:|||||||||||||:|||||| |||||| ||||||||| |
Db        876 NADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEK 935

Qy        944 GERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGP 1002
              |||||||||||||||::|||||||:: ||:|:|| :|||||||||||::||| :| :
Db        936 GERLPQPPICTIDVYIVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKL 995

Qy       1003 ASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDP---------------------- 1041
               || ||  |:::||:::|:  |::|||||||  |  |  | |
Db        996 PSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTSRTRIDSNRSEIGHSPPP 1054

Qy       1042 --APGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMG 1099
                 | :|    ::    :|: |   : :    :     | :|:|  :|| ::||     |
Db       1055 AYTPMSGSQFVYQDGGFATQQG---MPMPYTATTSTIPEAPVA--QGATAEMFDDSCCNG 1109

Qy       1100 AAKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVR 1152
                :   : | |||| ||||  |        : :||: |:   |: ||:|  :
Db       1110 TLRKPVVPHVQEDSSTQRYSADPTVFAPERNPRAELDEEGYMTPMHDKPKQEYLNPVE-- 1167

Qy       1153 PQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQ 1212
                    | |  : |          ||| ::      |::|||| :   |
Db       1168 -------ENPFVSRR--------------KNGDLQ-------ALDNPEYHSASSG---- 1194

Qy       1213 PHPPPA---------------------------------FSPAFDNLYYWDQDPPERGA 1238
```

```
Db       1195 --PPKAEDEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPDYWNHSLPPRST 1252

Qy       1239 --PPSTFKGTPT--------------AENPEYL 1255
Db       1253 LQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 8
XMRK_XIPMA
ID   XMRK_XIPMA     STANDARD;      PRT;  1167 AA.
AC   P13388;
DT   01-JAN-1990 (Rel. 13, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN   XMRK OR TU.
OS   Xiphophorus maculatus (Southern platyfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8083;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90015140; PubMed=2797166;
RA   Wittbrodt J., Adam D., Malitschek B., Maueler W., Raulf F.,
RA   Telling A., Robertson S.M., Schartl M.;
RT   "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT   inducing Tu locus in Xiphophorus.";
RL   Nature 341:415-421(1989).
RN   [2]
RP   REVISION TO 515.
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16891; CAA34770.2; -.
DR   PIR; S06142; S06142.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT   SIGNAL        1     25
FT   CHAIN        26   1167       MELANOMA RECEPTOR PROTEIN-TYROSINE
FT                                KINASE.
FT   DOMAIN        26    642       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     643    665       POTENTIAL.
FT   DOMAIN       666   1167       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       710    977       PROTEIN KINASE.
FT   NP_BIND      716    724       ATP (BY SIMILARITY).
FT   BINDING      743    743       ATP (BY SIMILARITY).
FT   ACT_SITE     835    835       BY SIMILARITY.
FT   DISULFID     195    204       BY SIMILARITY.
FT   DISULFID     199    212       BY SIMILARITY.
FT   DISULFID     220    228       BY SIMILARITY.
FT   DISULFID     224    236       BY SIMILARITY.
FT   DISULFID     237    245       BY SIMILARITY.
FT   DISULFID     241    253       BY SIMILARITY.
FT   DISULFID     256    265       BY SIMILARITY.
FT   DISULFID     269    296       BY SIMILARITY.
FT   DISULFID     300    311       BY SIMILARITY.
FT   DISULFID     315    330       BY SIMILARITY.
FT   DISULFID     333    337       BY SIMILARITY.
FT   DISULFID     504    513       BY SIMILARITY.
FT   DISULFID     508    521       BY SIMILARITY.
FT   DISULFID     524    533       BY SIMILARITY.
FT   DISULFID     537    553       BY SIMILARITY.
FT   DISULFID     556    569       BY SIMILARITY.
FT   DISULFID     560    577       BY SIMILARITY.
FT   DISULFID     593    615       BY SIMILARITY.
FT   DISULFID     618    626       BY SIMILARITY.
FT   DISULFID     622    634       BY SIMILARITY.
FT   CARBOHYD     114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1167 AA;  129934 MW;  4793E0749DC1D55A CRC64;

  Query Match             38.1%;  Score 2608.5;  DB 1;  Length 1167;
  Best Local Similarity   44.2%;  Pred. No. 1.5e-132;
  Matches  564;  Conservative  167;  Mismatches  396;  Indels  149;  Gaps   31;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
```

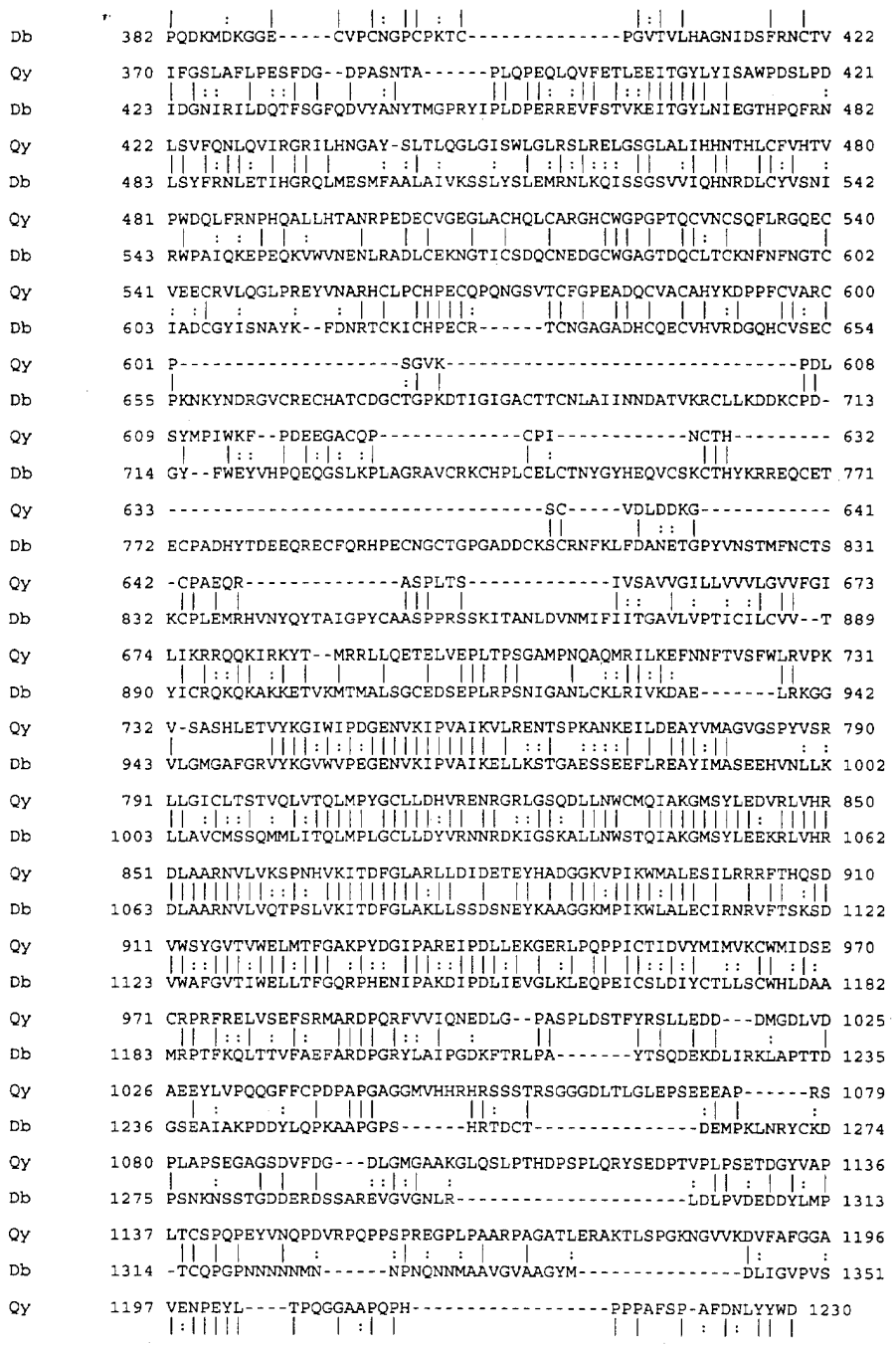
```
Db        417 NMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRY 476

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536

Db        477 ANTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCVSCLHVDR 529

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFC 596

Db        530 GGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKSAHFQDGPQC 589

Qy        597 VARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVS 656

Db        590 IPRCPHGILGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGD-IVSHSSLAVG 647

Qy        657 AVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716

Db        648 LVSGLLITVIVALLIVVLLRRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQAFLRILK 706

Qy        717 --EFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIL 774

Db        707 ETEFKKDRV--------LGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSPKVNQEVL 758

Qy        775 DEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQ 834

Db        759 DEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCVQ 818

Qy        835 IAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKW 894

Db        819 IAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPIKW 878

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954

Db        879 MALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPICT 938

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014

Db        939 IEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSLSDRRLFSRL 995

Qy       1015 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1074

Db        996 LSSDD--DVVDADEYLLPYKRI---------------------NRQGS------------ 1020

Qy       1075 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDGY 1133

Db       1021 ----EPCIPPTGH--------------------PVRENSITLRNISDPTQNALEKDLDGH 1056

Qy       1134 VAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKTLS 1180

Db       1057 ----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYLN 1106

Qy       1181 PGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPP 1240

Db       1107 TNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGALT 1141

Qy       1241 STFKGTPTAENPEYLG 1256

Db       1142 GNGMFLPAAENLEYLG 1157


RESULT 9
ERB3_HUMAN
ID   ERB3_HUMAN     STANDARD;      PRT;  1342 AA.
AC   P21860;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN   ERBB3 OR HER3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90083234; PubMed=2687875;
RA   Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT   "Isolation and characterization of ERBB3, a third member of the
RT   ERBB/epidermal growth factor receptor family: evidence for
RT   overexpression in a subset of human mammary tumors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90311312; PubMed=2164210;
RA   Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA   Todaro G.J., Shoyab M.;
RT   "Molecular cloning and expression of an additional epidermal growth
RT   factor receptor-related gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN   [3]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RC   TISSUE=Placenta;
RX   MEDLINE=93282822; PubMed=7685162;
RA   Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT   "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT   tyrosine kinase.";
RL   Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC       SECRETED (SHORT FORM).
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC       SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC   -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29366; AAA35790.1; -.
DR   EMBL; M34309; AAA35979.1; -.
DR   EMBL; S61953; AAB26935.1; -.
DR   PIR; A36223; A36223.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3431; ERBB3.
DR   MIM; 190151; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
```

```
KW   Alternative splicing.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1342       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN       20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    644    664       POTENTIAL.
FT   DOMAIN      665   1342       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      709    966       PROTEIN KINASE.
FT   NP_BIND     715    723       ATP (BY SIMILARITY).
FT   BINDING     742    742       ATP (BY SIMILARITY).
FT   ACT_SITE    834    834       BY SIMILARITY.
FT   DISULFID    186    194       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    552    565       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    141    183       EILSGGVYIEKNDKLCHMDTIDWRDIVRDRDAEIVVKDNGR
FT                                SC -> GQFPMVPSGLTPQPAQDWYLLDDDPRLLTLSASSK
FT                                VPVTLAAV (IN SHORT ISOFORM).
FT   VARSPLIC    184   1342       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    560    560       E -> G (IN REF. 2).
FT   CONFLICT   1064   1064       E -> G (IN REF. 2).
SQ   SEQUENCE   1342 AA;  148097 MW;  7201E7F66CA374BD CRC64;

  Query Match             34.2%;  Score 2343.5;  DB 1;  Length 1342;
  Best Local Similarity   39.7%;  Pred. No. 2.9e-118;
  Matches  524;  Conservative  195;  Mismatches  459;  Indels  143;  Gaps   36;

Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
Db        238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFQLEPNPHTKYQYGGVCVASCPHNFV-V 296

Qy        307 DVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF--IGITELE- 363
Db        297 DQTSCVRACPPDKMEVD-KNGLKMCEPCGGLCPKACEGTG-----SGSRFQTVDSSNIDG 350

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
Db        351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVFRTVREITGYLNIQSWPPHMHNFS 410

Qy        424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        411 VFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNW 470

Qy        483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECV 541
Db        471 TKVLRGPTEERLDIKHNRPRRDCVAEGKVCDPLCSSGGCWGPGPGQCLSCRNYSRGGVCV 530

Qy        542 EECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCP 601
Db        531 THCNFLNGEPREFAHEAECFSCHPECQPMEGTATCNGSGSDTCAQCAHFRDGPHCVSSCP 590

Qy        602 SGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRA----SPLTSIVSA 657
Db        591 HGVLG--AKGPIYKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLIGKTHLTMALTV 648

Qy        658 VVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
Db        649 IAG--LVVIFMMLGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705

Qy        717 EFNNFTVSFWLRVPKVSASHL-ETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILD 775
Db        706 ETE-------LRKLKVLGSGVFGTVHKGVWIPEGESIKIPVCIKVIEDKSGRQSFQAVTD 758

Qy        776 EAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQI 835
Db        759 HMLAIGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVRQHRGALGPQLLLNWGVQI 818

Qy        836 AKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWM 895
Db        819 AKGMYYLEEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKWM 878

Qy        896 ALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTI 955
Db        879 ALESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEVPDLLEKGERLAQPQICTI 938

Qy        956 DVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPA---SPLDSTFYR 1012
Db        939 DVYMVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTN 997

Qy       1013 SLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP- 1071
Db        998 KKLEEVELEPELDLDLDLEAEED------------------NLATTTLGSALSLPVGTL 1038

Qy       1072 SEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHD-PSPLQRYSEDPTVPLP--- 1127
Db       1039 NRPRGSQSLLSPSSGY-MPMNQGNLGESCQESAVSGSSERCPRPVSLH------PMPRGC 1091

Qy       1128 ---SETDGYVA----------------PLTCSPQPE----YVNQPDVRPQPPSPREGP-- 1162
Db       1092 LASESSEGHVTGSEAELQEKVSMCRSRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPPGL 1151

Qy       1163 ---------LPAARPAGATLERAKTLSP-GKNGVV------KDVFAFGGAVENPEYLTPQ 1206
Db       1152 EEEDVNGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEEEDED--------EEYEYMNRR 1203

Qy       1207 GGAAPQPHPPPAFSPAFDNLYYWD------------QDPPERGAPPSTFKGTPTAENPEY 1254
Db       1204 RRHSP-PHPPRPSSLEELGYEYMDVGSDLSASLGSTQSCPLHPVPIMPTAGTTPDEDYEY 1262

Qy       1255 L 1255
Db       1263 M 1263
```

```
RESULT 10
ERB3_RAT
ID   ERB3_RAT       STANDARD;      PRT;  1339 AA.
AC   Q62799; Q62955;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3).
GN   ERBB3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96096535; PubMed=8522190;
RA   Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT   "Cloning of the rat ErbB3 cDNA and characterization of the
RT   recombinant protein.";
RL   Gene 165:279-284(1995).
RN   [2]
RP   REVISIONS TO 85; 513 AND 565.
RA   Hellyer N.J., Koland J.G.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 922-1097 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29339; AAC28498.2; -.
DR   EMBL; U52530; AAC53050.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1339       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN       20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    644    662       POTENTIAL.
FT   DOMAIN      663   1339       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      183    259       CYS-RICH.
FT   DOMAIN      707    964       PROTEIN KINASE.
FT   NP_BIND     713    721       ATP (BY SIMILARITY).
FT   BINDING     740    740       ATP (BY SIMILARITY).
FT   ACT_SITE    832    832       BY SIMILARITY.
FT   DISULFID    186    194       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1028   1028       L -> P (IN REF. 3).
SQ   SEQUENCE   1339 AA;  147545 MW;  0AA5F2402BBFDF1E CRC64;

  Query Match             33.1%;  Score 2271.5;  DB 1;  Length 1339;
  Best Local Similarity   40.0%;  Pred. No. 2.1e-114;
  Matches  517;  Conservative  171;  Mismatches  437;  Indels  167;  Gaps   37;

Qy          3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              |  ||     |  |  |:    |  || ||   | :    :     |  ||: |:|| ||
Db          7 LQVLC----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLYKLYEKCEVVMGN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:     || ||||| |:|| ||||:| |:   :||  ||:|||||::  :|: |:
Db         63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLPNLRVVRGTQVYDGKFAIFVM-- 120

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 ||  |   |   ||:|:  |||| ||| |:| :||: ||| |:||
Db        121 ---LNYNT----NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRVR--- 170

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
               ::  |    | ||  :||| |||   :||| ||:|:||  |  || || |  |||:
Db        171 GAEIVVKNNGANCPPCHEVCKG-RCWGPGPDDCQILTKTICAPQCNGRCFGPNPNQCCHD 229

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              :|| ||:||: :|| ||  || || |   ||  : ||  ||:  ||| :| :|  || :
Db        230 ECAGGCSGPQDTDCFACRRFNDSGACVPRCPEPLVYNKLTFQLEPNPHTKYQYGGVCVAS 289

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGL--GMQYIKANSKF 356
              ||:|::  |   |   ||    ||  : | : || |   | : | |   | :|   :|
Db        290 CPHNFV-VDQTFCVRACPPDKMEVD-KHGLKMCEPCGGLCPKACEGTGSGSRYQTVDSSN 347

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
```

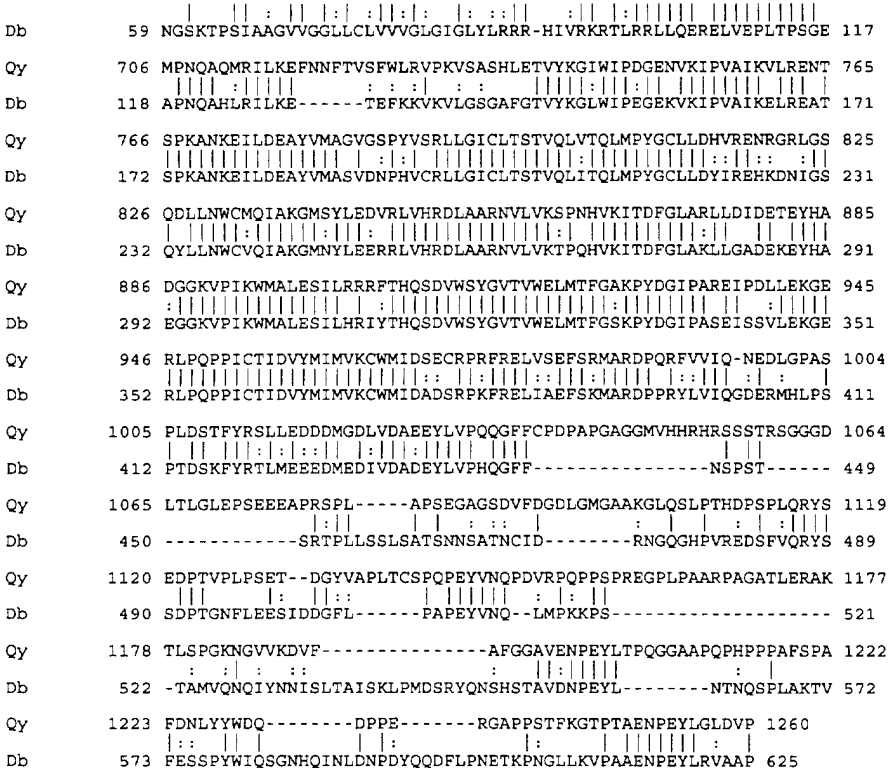
```
Db        348 ID----GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLNVFRTVREITGYLNIQSWP 403

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLC 475

Db        404 PHMHNFSVFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLC 463

Qy        476 FVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQF 534

Db        464 YHHSLNWTRLLRGPSEERLDIKYDRPLGECLAEGKVCDPLCSSGGCWGPGPGQCLSCRNY 523

Qy        535 LRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPP 594

Db        524 SREGVCVTHCNFLQGEPREFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAHFRDGP 583

Qy        595 FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSC--VDLDDKGCPAEQRASPLT 652

Db        584 HCVNSCPHGILG--AKGPIYKYPDAQNECRPCHENCTQGCNGPELQDCLGQAEVLMSKPH 641

Qy        653 SIVSAVVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGAMPNQAQ 711

Db        642 LVIAVTVG--LAVILMILGGSFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVL 698

Qy        712 MRILKEFNNFTVSFWLRVPKVSASHL-ETVYKGIWIPDGENVKIPVAIKVLRENTSPKAN 770

Db        699 ARIFKETE-------LRKLKVLGSGVFGTVHKGIWIPEGESIKIPVCIKVIEDKSGRQSF 751

Qy        771 KEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLN 830

Db        752 QAVTDHMLAVGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVKQHRETLGPQLLLN 811

Qy        831 WCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKV 890

Db        812 WGVQIAKGMYYLEEHSMVHRDLALRNVMLKSPSQVQVADFGVADLLPPDDKQLLHSEAKT 871

Qy        891 PIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQP 950

Db        872 PIKWMALESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQP 931

Qy        951 PICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTF 1010

Db        932 QICTIDVYMVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAA 988

Qy       1011 YRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLE 1070

Db        989 EPSVLTTKEL-----QEAELEPEL----------------------------DLDLDLE 1014

Qy       1071 PSEE--------------------EAPRSPLAPSEG--------AGSDVFDGDLGMGAAK 1102

Db       1015 AEEEGLATSLGSALSLPTGTLTRPRGSQSLLSPSSGYMPMNQSSLGEACLDSAVLGGREQ 1074

Qy       1103 GLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV----APL-----TC-------SPQPE-- 1144

Db       1075 FSRPISLH-PIPRGR-------PASESSEGHVTGSEAELQEKVSVCRSRSRSRSPRPRGD 1126

Qy       1145 --YVNQPDVRPQPPSPREGP-----------LPAARPAGATLERAKTLSP-GKNGVV--- 1187

Db       1127 SAYHSQRHSLLTPVTPLSPPGLEEEDGNGYVMPDTHLRGASSSREGTLSSVGLSSVLGTE 1186

Qy       1188 ---KDVFAFGGAVENPEYLTPQGGAAPQPHPP 1216

Db       1187 EEDED--------EEYEYMNRKRRGSP-PRPP 1209


RESULT 11
EGFR_DROME
ID   EGFR_DROME     STANDARD;      PRT;  1426 AA.
AC   P04412; O61601; Q9W2G0; P81868;
DT   13-AUG-1987 (Rel. 05, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE   (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN   EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX   MEDLINE=94350209; PubMed=8070664;
RA   Clifford R., Schupbach T.;
RT   "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT   that several genetically defined classes of alleles cluster in
RT   subdomains of the receptor protein.";
RL   Genetics 137:531-550(1994).
RN   [2]
RP   REVISIONS.
RA   Clifford R., Schupbach T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=85124611; PubMed=2982499;
RA   Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT   "The Drosophila EGF receptor gene homolog: conservation of both
RT   hormone binding and kinase domains.";
RL   Cell 40:599-607(1985).
RN   [4]
RP   SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC   STRAIN=Oregon-R; TISSUE=Embryo;
RX   MEDLINE=87002474; PubMed=3093080;
RA   Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT   "Alternative 5' exons and tissue-specific expression of the
RT   Drosophila EGF receptor homolog transcripts.";
RL   Cell 46:1091-1101(1986).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=99102120; PubMed=9882502;
RA   Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT   "Several levels of EGF receptor signaling during photoreceptor
RT   specification in wild-type, Ellipse, and null mutant Drosophila.";
RL   Dev. Biol. 205:129-144(1999).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```

```
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [7]
RP   SEQUENCE OF 959-1078 FROM N.A.
RC   STRAIN=Daekwanryeong;
RX   MEDLINE=85137938; PubMed=2983232;
RA   Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT   "A Drosophila genomic sequence with homology to human epidermal
RT   growth factor receptor.";
RL   Nature 314:178-180(1985).
RN   [8]
RP   SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=92038942; PubMed=1936959;
RA   Raz E., Schejter E.D., Shilo B.Z.;
RT   "Interallelic complementation among DER/flb alleles: implications for
RT   the mechanism of signal transduction by receptor-tyrosine kinases.";
RL   Genetics 129:191-201(1991).
RN   [9]
RP   REVIEW.
RX   MEDLINE=97248481; PubMed=9094709;
RA   Perrimon N., Perkins L.A.;
RT   "There must be 50 ways to rule the signal: the case of the Drosophila
RT   EGF receptor.";
RL   Cell 89:13-16(1997).
CC   -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC       WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC       MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC       CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC       DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC       POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC       ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC       AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC       SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC       CUTICLE.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC       MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC       PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC       TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC       UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC       TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC       EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC       IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC       POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC       AND THORACIC AND ABDOMINAL GANGLIA.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF052754; AAC08536.1; -.
DR   EMBL; AF052753; AAC08536.1; JOINED.
DR   EMBL; AF052754; AAC08535.1; -.
DR   EMBL; AF052752; AAC08535.1; JOINED.
DR   EMBL; K03054; AAA51462.1; -.
DR   EMBL; K03417; AAA51460.1; -.
DR   EMBL; K03416; AAA50965.1; -.
DR   EMBL; K03418; AAA51461.1; -.
DR   EMBL; AF109077; AAD26134.1; -.
DR   EMBL; AF109078; AAD26132.1; -.
DR   EMBL; AF109082; AAD26132.1; JOINED.
DR   EMBL; AF109078; AAD26133.1; -.
DR   EMBL; AF109084; AAD26133.1; JOINED.
DR   EMBL; AF109079; AAD26130.1; -.
DR   EMBL; AF109081; AAD26130.1; JOINED.
DR   EMBL; AF109079; AAD26131.1; -.
DR   EMBL; AF109083; AAD26131.1; JOINED.
DR   EMBL; AF109080; AAD26135.1; -.
DR   EMBL; AE003454; AAF46732.1; -.
DR   EMBL; X02293; CAA26157.1; -.
DR   EMBL; X78920; CAA55523.1; -.
DR   EMBL; X78918; CAA55521.1; -.
DR   EMBL; X78919; CAA55522.1; -.
DR   PIR; A00640; GQFFE.
DR   HSSP; P11362; 1FGK.
DR   FlyBase; FBgn0003731; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 7.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW   Developmental protein.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1426       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    868       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    869    889       POTENTIAL.
FT   DOMAIN      890   1426       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      938   1198       PROTEIN KINASE.
FT   NP_BIND     944    952       ATP (BY SIMILARITY).
FT   BINDING     971    971       ATP (BY SIMILARITY).
FT   ACT_SITE   1063   1063       BY SIMILARITY.
FT   MOD_RES     902    902       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

  Query Match             27.4%;  Score 1879;  DB 1;  Length 1426;
  Best Local Similarity   32.0%;  Pred. No. 2.4e-93;
  Matches  460;  Conservative  183;  Mismatches  436;  Indels  358;  Gaps   44;

Qy         24 QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT-NASLSFLQDIQEVQG 82
              ::| ||  :| :|:: | |   ||  |  |  | |||:||:||  |  |||| :|:|| |
Db        100 KICIGTKSRLSVPSNKEHHYRNLRDRYTNCTYVDGNLKLTWLPNENLDLSFLDNIREVTG 159

Qy         83 YVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGL 137
              |:||:|  |::|   :|:|:||  ||     |: ||| |             :    :
Db        160 YILISHVDVKKVVFPKLQIIRGRTLFSLSVEEEKYALFV---------------TYSKM 203

Qy        138 RELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSP 197
               |::  | ::| | |    |  ||:  || | :|           |   | |  |
Db        204 YTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPKCHE 263

Qy        198 MCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDCLAC 255
               |    ||||  ::||  ::  |:  ||  || || | :|||  || |||||    ||:||
Db        264 SCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDCIAC 322

Qy        256 LHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVC 315
              :|    : :  || :  ||  |:    ||||:| :||:|| ||  :|  | |:|   |
Db        323 KNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACVRSC 381

Qy        316 PLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAGCKK 369
```



```
Db        382 PQDKMDKGGE-----CVPCNGPCPKTC--------------PGVTVLHAGNIDSFRNCTV 422

Qy        370 IFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDSLPD 421

Db        423 IDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQFRN 482

Qy        422 LSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db        483 LSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYVSNI 542

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db        543 RWPAIQKEPEQKVWVNENLRADLCEKNGTICSDQCNEDGCWGAGTDQCLTCKNFNFNGTC 602

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db        603 IADCGYISNAYK--FDNRTCKICHPECR------TCNGAGADHCQECVHVRDGQHCVSEC 654

Qy        601 P--------------------SGVK--------------------------------PDL 608

Db        655 PKNKYNDRGVCRECHATCDGCTGPKDTIGIGACTTCNLAIINNDATVKRCLLKDDKCPD- 713

Qy        609 SYMPIWKF--PDEEGACQP-------------CPI------------NCTH--------- 632

Db        714 GY--FWEYVHPQEQGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCET 771

Qy        633 ------------------------------------SC-----VDLDDKG----------- 641

Db        772 ECPADHYTDEEQRECFQRHPECNGCTGPGADDCKSCRNFKLFDANETGPYVNSTMFNCTS 831

Qy        642 -CPAEQR--------------ASPLTS-------------IVSAVVGILLVVVLGVVFGI 673

Db        832 KCPLEMRHVNYQYTAIGPYCAASPPRSSKITANLDVNMIFIITGAVLVPTICILCVV--T 889

Qy        674 LIKRRQQKIRKYT--MRRLLQETELVEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPK 731

Db        890 YICRQKQKAKKETVKMTMALSGCEDSEPLRPSNIGANLCKLRIVKDAE-------LRKGG 942

Qy        732 V-SASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSR 790

Db        943 VLGMGAFGRVYKGVWVPEGENVKIPVAIKELLKSTGAESSEEFLREAYIMASEEHVNLLK 1002

Qy        791 LLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHR 850

Db       1003 LLAVCMSSQMMLITQLMPLGCLLDYVRNNRDKIGSKALLNWSTQIAKGMSYLEEKRLVHR 1062

Qy        851 DLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSD 910

Db       1063 DLAARNVLVQTPSLVKITDFGLAKLLSSDSNEYKAAGGKMPIKWLALECIRNRVFTSKSD 1122

Qy        911 VWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSE 970

Db       1123 VWAFGVTIWELLTFGQRPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCWHLDAA 1182

Qy        971 CRPRFRELVSEFSRMARDPQRFVVIQNEDLG--PASPLDSTFYRSLLEDD---DMGDLVD 1025

Db       1183 MRPTFKQLTTVFAEFARDPGRYLAIPGDKFTRLPA-------YTSQDEKDLIRKLAPTTD 1235

Qy       1026 AEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAP------RS 1079

Db       1236 GSEAIAKPDDYLQPKAAPGPS------HRTDCT--------------DEMPKLNRYCKD 1274

Qy       1080 PLAPSEGAGSDVFDG---DLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136

Db       1275 PSNKNSSTGDDERDSSAREVGVGNLR--------------------LDLPVDEDDYLMP 1313

Qy       1137 LTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGA 1196

Db       1314 -TCQPGPNNNNNMN------NPNQNNMAAVGVAAGYM--------------DLIGVPVS 1351

Qy       1197 VENPEYL----TPQGGAAPQPH-----------------PPPAFSP-AFDNLYYWD 1230

Db       1352 VDNPEYLLNAQTLGVGESPIPTQTIGIPVMGGPGTMEVKVPMPGSEPTSSDHEYYND 1408
```

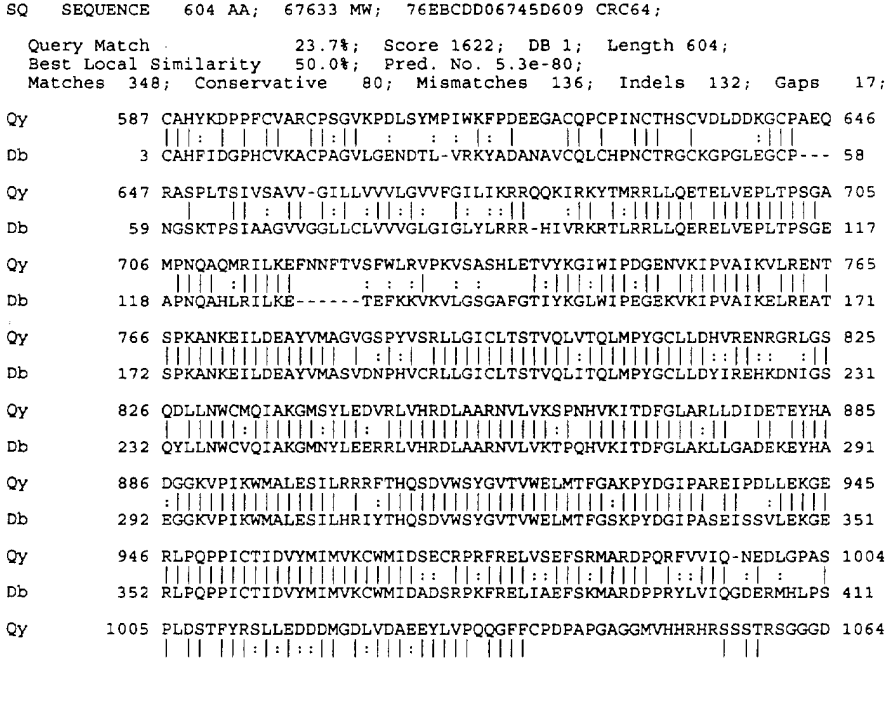
```
RESULT 12
ERBB_ALV
ID   ERBB_ALV       STANDARD;      PRT;   634 AA.
AC   P00534;
DT   21-JUL-1986 (Rel. 01, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian leukosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11864;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85228222; PubMed=2988784;
RA   Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA   Crittenden L.B., Raines M.A., Kung H.-J.;
RT   "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT   processing and promoter insertion result in expression of an
RT   amino-truncated EGF receptor.";
RL   Cell 41:719-726(1985).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC       IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC       AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC       THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC       C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC       PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10066; AAA48763.1; ALT_INIT.
DR   PIR; A00643; TVCHLV.
DR   PIR; B00643; TVFVLV.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
SQ   SEQUENCE   634 AA;  70891 MW;  E705E33A0BE01FCC CRC64;

  Query Match             24.3%;  Score 1668.5;  DB 1;  Length 634;
  Best Local Similarity   50.2%;  Pred. No. 1.8e-82;
  Matches  358;  Conservative   83;  Mismatches  143;  Indels  129;  Gaps   18;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646

Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705
```



```
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENT 765
              |||| :|||||      : : :|  : :    |||||:|||:|| ||||||||  ||| |
Db        118 APNQAHLRILKE------TEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREAT 171

Qy        766 SPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGS 825
              |||||||||||||||| | :|:| ||||||||||||:|||||||||||::||::  :||
Db        172 SPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGS 231

Qy        826 QDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHA 885
              | |||||:||||||:|||: |||||||||||||||:| ||||||||||:||  || ||||
Db        232 QYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHA 291

Qy        886 DGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGE 945
              :|||||||||||||| | :|||||||||||||||||:|||||||| ||  :|||||
Db        292 EGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGE 351

Qy        946 RLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPAS 1004
              |||||||||||||||||||||:: ||:||||::|||:|||||  |::||| :| :   |
Db        352 RLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPS 411

Qy       1005 PLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGD 1064
              | || |||:|:|::|| |:|||:|||||  ||||                  | ||
Db        412 PTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------ 449

Qy       1065 LTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYS 1119
                          |:||    | |  : ::  |        :  |  |  : | :||||
Db        450 ------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYS 489

Qy       1120 EDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAK 1177
               |||      |:  ||::      | ||||||  : |: ||
Db        490 SDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS-------------------- 521

Qy       1178 TLSPGKNGVVKDVF---------------AFGGAVENPEYLTPQGGAAPQPHPPPAPSPA 1222
                :  :| :  ::                :   ||:|||||          :  |
Db        522 -TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTV 572

Qy       1223 FDNLYYWDQ--------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1260
              |::  || |        | |:          |:     | |||||| :  |
Db        573 FESSPYWIQSGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 625
```

RESULT 13
ERBB_AVIER
ID   ERBB_AVIER   STANDARD;   PRT;   604 AA.
AC   P00535;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ES4).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=79685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H;
RX   MEDLINE=84026539; PubMed=6313229;
RA   Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT   "The erbB gene of avian erythroblastosis virus is a member of the src
RT   gene family.";
RL   Cell 35:71-78(1983).
RN   [2]
RP   SEQUENCE OF 1-152 FROM N.A.
RX   MEDLINE=84223957; PubMed=6328658;
RA   Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA   Saule S., Martin P., Stehelin D.;
RT   "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT   new type of oncogene.";
RL   Science 224:1456-1459(1984).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC   tyrosine phosphate.
CC   -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC   ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC   IN CHICKENS.
CC   -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC   RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; K02006; AAA42394.1; ALT_INIT.
DR   EMBL; K01216; AAA42400.1; -.
DR   PIR; A00644; TVYUH.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN   132   399   PROTEIN KINASE.
FT   NP_BIND   138   146   ATP (BY SIMILARITY).
FT   BINDING   165   165   ATP (BY SIMILARITY).
FT   ACT_SITE   257   257   BY SIMILARITY.
FT   CONFLICT   29   29   R -> W (IN REF. 2).
FT   CONFLICT   140   140   S -> F (IN REF. 2).
FT   CONFLICT   146   146   I -> V (IN REF. 2).
SQ   SEQUENCE   604 AA; 67633 MW; 76EBCDD06745D609 CRC64;

  Query Match   23.7%; Score 1622; DB 1; Length 604;
  Best Local Similarity   50.0%; Pred. No. 5.3e-80;
  Matches 348; Conservative   80; Mismatches 136; Indels 132; Gaps   17;



```
Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
              |||: | | ||  ||:||  :   : : |: |    || |  |||  |     :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705
                |   || : || |:| :||:|:  |: ::||   :|| |:|||||  ||||||||||
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENT 765
              |||| :|||||      : : :|  : :    |:|||:|||:|| ||||||||  ||| |
Db        118 APNQAHLRILKE------TEFKKVKVLGSGAFGTIYKGLWIPEGEKVKIPVAIKELREAT 171

Qy        766 SPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGS 825
              |||||||||||||||| | :|:| ||||||||||||:|||||||||||::||::  :||
Db        172 SPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGS 231

Qy        826 QDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHA 885
              | |||||:||||||:|||: |||||||||||||||:| ||||||||||:||  || ||||
Db        232 QYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHA 291

Qy        886 DGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGE 945
              :|||||||||||||| | :|||||||||||||||||:|||||||| ||  :|||||
Db        292 EGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGE 351

Qy        946 RLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPAS 1004
              |||||||||||||||||||||:: ||:||||::|||:|||||  |::||| :| :   |
Db        352 RLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPS 411

Qy       1005 PLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGD 1064
              | || |||:|:|::|| |:|||:|||||  ||||                  | ||
```

```
Db       412 PTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------ 449

Qy      1065 LTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYS 1119
                          |:||     | |  : ::  |         :  |  |  : | :||||
Db       450 ------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYS 489

Qy      1120 EDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAK 1177
               |||      |:  ||::      | ||||||  : |: ||
Db       490 SDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPSTAM----------------- 524

Qy      1178 TLSPGKNGVVKDVFAF-------------------GGAVENPEYLTPQGGAAPQPHPPPA 1218
                        |   :: |                    ||:|||||          :  |
Db       525 --------VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPL 568

Qy      1219 FSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254
                |::  || |    :             :||:|
Db       569 AKTVFESSPYWIQSGNHQ----------INLDNPDY 594


RESULT 14
EGFR_CHICK
ID   EGFR_CHICK     STANDARD;      PRT;   703 AA.
AC   P13387;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE   (Fragment).
GN   EGFR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88261272; PubMed=3260329;
RA   Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA   Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT   "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT   expression in mouse cells, and differential binding of EGF and
RT   transforming growth factor alpha.";
RL   Mol. Cell. Biol. 8:1970-1978(1988).
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20386; AAA48760.1; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 4.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1    30
FT   CHAIN        31   >703       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31   654        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655   667        POTENTIAL.
FT   DOMAIN      668   >703       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    197   206        BY SIMILARITY.
FT   DISULFID    201   214        BY SIMILARITY.
FT   DISULFID    222   230        BY SIMILARITY.
FT   DISULFID    226   238        BY SIMILARITY.
FT   DISULFID    239   247        BY SIMILARITY.
FT   DISULFID    243   255        BY SIMILARITY.
FT   DISULFID    258   267        BY SIMILARITY.
FT   DISULFID    271   298        BY SIMILARITY.
FT   DISULFID    302   314        BY SIMILARITY.
FT   DISULFID    318   333        BY SIMILARITY.
FT   DISULFID    336   340        BY SIMILARITY.
FT   DISULFID    513   522        BY SIMILARITY.
FT   DISULFID    517   530        BY SIMILARITY.
FT   DISULFID    533   542        BY SIMILARITY.
FT   DISULFID    546   562        BY SIMILARITY.
FT   DISULFID    565   581        BY SIMILARITY.
FT   DISULFID    569   589        BY SIMILARITY.
FT   DISULFID    592   601        BY SIMILARITY.
FT   DISULFID    605   627        BY SIMILARITY.
FT   DISULFID    630   638        BY SIMILARITY.
FT   DISULFID    634   646        BY SIMILARITY.
FT   CARBOHYD    134   134        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190   190        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    200   200        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    359   359        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368   368        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    420   420        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573   573        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578   578        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613   613        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    633   633        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    648   648        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     703   703
SQ   SEQUENCE   703 AA;  77427 MW;  AFF2DE11B735A690 CRC64;

  Query Match           23.3%;  Score 1595;  DB 1;  Length 703;
  Best Local Similarity   44.6%;  Pred. No. 1.8e-78;
  Matches  316;  Conservative  111;  Mismatches  251;  Indels   30;  Gaps   14;

Qy         8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
             |   :| ||  | |          :|| ||: ||      | |    |: :|  |:||  |||
Db        13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy        62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
               :||:  |  |:||: |||| |||||| | |  :||: |:|:||  |:::::|||| |
Db        73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy       122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
              :| |          ||||| :: |:||| ||| |  ||:||  ||:|| ||   : :  |
Db       133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy       182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
             |::|  :   :|  | | |     |||   ::||:||: :||  |: ||:| :|:|||| |
Db       183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy       240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
             |||||||: ||||||  |      |:  || || ||  |::   ||||:|:|||:|| |
Db       243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy       300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
             |:|| || |||    |     ||  |:| ::|:||   |::|| |:|:  :|     |
Db       303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy       360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
             | :: |  | || | :: || :| ||  : | || |::| || |::||:|:| | ||||:
```

```
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478

Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538

Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595

Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIV 655

Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 SAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTP 702

Db        657 AGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTP 703


RESULT 15
ERBB_AVIEU
ID   ERBB_AVIEU     STANDARD;      PRT;   540 AA.
AC   P11273;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ts167).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=103898;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87064458; PubMed=2878364;
RA   Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT   "A single amino acid substitution in v-erbB confers a thermolabile
RT   phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT   cells.";
RL   Mol. Cell. Biol. 6:1751-1759(1986).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13179; AAA42401.1; -.
DR   PIR; A25231; TVFVEB.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   VARIANT     270    270       H -> D (IN THERMOLABILE V-ERBB).

SQ   SEQUENCE   540 AA;  60412 MW;  5B53297AA068B65D CRC64;

  Query Match             22.6%;  Score 1552;  DB 1;  Length 540;
  Best Local Similarity   52.8%;  Pred. No. 2.6e-76;
  Matches  330;  Conservative   72;  Mismatches  127;  Indels   96;  Gaps   15;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646

Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705

Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENT 765

Db        118 APNQAHLRILKE------TEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELREAT 171

Qy        766 SPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGS 825

Db        172 SPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGS 231

Qy        826 QDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHA 885

Db        232 QYLLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAKQLGADEKEYHA 291

Qy        886 DGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGE 945

Db        292 EGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGE 351

Qy        946 RLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPAS 1004

Db        352 RLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPS 411

Qy       1005 PLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGD 1064

Db        412 PTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST------ 449

Qy       1065 LTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYS 1119

Db        450 ------------SRTPLLSSLSATSNNSATNCIDRNGG------------H--------- 476

Qy       1120 EDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERAKT 1178

Db        477 -------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAISKLPI 521

Qy       1179 LSPGKNGVVKDVFAFGGAVENPEYL 1203

Db        522 DSRYQN-------SHSTAVDNPEYL 539


Search completed: July 22, 2003, 09:18:43
Job time : 22.6928 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 47.0485 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-695-709-14
Perfect score: 6853
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6130	89.4	1259	6	O18735	O18735 canis famil
2	3063	44.7	1209	11	Q9QX70	Q9qx70 rattus norv
3	3034	44.3	1210	11	Q9EP98	Q9ep98 mus musculu
4	2650	38.7	1165	13	Q9YH40	Q9yh40 xiphophorus
5	2612.5	38.1	1137	13	Q9W6F6	Q9w6f6 gallus gall
6	2226	32.5	1328	13	P79754	P79754 fugu rubrip
7	1959.5	28.6	1433	5	Q9BIH9	Q9bih9 anopheles g
8	1871	27.3	419	4	Q9UK79	Q9uk79 homo sapien
9	1739	25.4	367	11	Q8R2X1	Q8r2x1 mus musculu
10	1697.5	24.8	412	4	Q8WYV0	Q8wyv0 homo sapien
11	1639	23.9	729	15	Q86712	Q86712 avian rous-
12	1637	23.9	567	15	Q86714	Q86714 avian rous-
13	1575.5	23.0	962	15	Q64895	Q64895 avian eryth
14	1567	22.9	545	15	Q85468	Q85468 avian eryth
15	1506.5	22.0	655	11	Q9WVF5	Q9wvf5 mus musculu
16	1490.5	21.7	643	11	Q9ERV6	Q9erv6 mus musculu
17	1204	17.6	1193	5	Q9Y1X8	Q9y1x8 ephydatia f
18	1139.5	16.6	1368	5	Q23821	Q23821 caenorhabdi
19	1129	16.5	1717	5	Q26566	Q26566 schistosoma
20	1126	16.4	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.6	478	11	Q9ESE0	Q9ese0 rattus norv
22	942.5	13.8	599	13	Q9PSH2	Q9psh2 gallus gall
23	906	13.2	165	4	Q14256	Q14256 homo sapien
24	887	12.9	176	11	Q923V5	Q923v5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8szw1 drosophila
27	754.5	11.0	311	13	Q99162	Q99162 xiphophorus
28	734	10.7	331	4	Q9BUD7	Q9bud7 homo sapien
29	723	10.6	149	6	Q9BG66	Q9bg66 oryctolagus
30	705.5	10.3	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
31	687	10.0	1671	5	Q9NJV5	Q9njv5 biomphalari
32	653.5	9.5	1368	13	Q8UW85	Q8uw85 paralichthy
33	645	9.4	1418	13	O93457	O93457 scophthalmu
34	640.5	9.3	1369	13	Q8UW86	Q8uw86 paralichthy
35	628	9.2	1358	13	O73798	O73798 xenopus lae
36	626.5	9.1	1472	5	Q9U5A8	Q9u5a8 bombyx mori
37	611	8.9	1412	13	Q8UW84	Q8uw84 paralichthy
38	601	8.8	1245	13	Q9YGH8	Q9ygh8 scophthalmu
39	597.5	8.7	1418	13	Q8UW83	Q8uw83 paralichthy
40	587.5	8.6	2144	5	Q9VD94	Q9vd94 drosophila
41	577	8.4	1371	11	Q9QVW4	Q9qvw4 rattus sp.
42	572.5	8.4	1091	4	Q9UMQ4	Q9umq4 homo sapien
43	564	8.2	987	11	Q91YM0	Q91ym0 mus musculu
44	562.5	8.2	1072	4	Q9BTB0	Q9btb0 homo sapien
45	561	8.2	935	4	Q96L35	Q96l35 homo sapien

ALIGNMENTS

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErbB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB008451; BAA23127.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

```
  Query Match             89.4%;  Score 6130;  DB 6;  Length 1259;
  Best Local Similarity   89.6%;  Pred. No. 0;
  Matches 1135;  Conservative   43;  Mismatches   75;  Indels   14;  Gaps    4;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL  60
              ||||| |||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGLLLALLPSGAAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL  60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG  120
              |||||| ||||||||||||||||||||:||||:|||||||||||||||||||||||||||
Db         61 ELTYLPANASLSFLQDIQEVQGYVLIAHSQVRQIPLQRLRIVRGTQLFEDNYALAVLDNG  120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA  180
              |||    |  ||: |||||||||||||||||||||:||||:||||||||:||||||||
Db        121 DPLEGGIPAPGAAQGGLRELQLRSLTEILKGGVLIQRSPQLCHQDTILWKDVFHKNNQLA  180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC  240
              |||||||  || |||| || : ||| || ||||||||||||||||||| |||||||||
Db        181 LTLIDTNRFSACPPCSPACKDAHCWGASSGDCQSLTRTVCAGGCARCKGPQPTDCCHEQC  240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP  300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT  360
              |||||||||||||||:|||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN  360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP  420
              ||||||||||||||||||||||||||||||||:||| ||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLRVFEALEEITGYLYISAWPDSLP  420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV  480
              :|||||:|||||:||:|||||||||||||||||||||||||||| |  |||||||
Db        421 NLSVFQNLRVIRGRVLHDGAYSLTLQGLGISWLGLRSLRELGSGLALIHRNARLCFVHTV  480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC  540
              |||||||||||||:|||||:|||||||||:  || ||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHSANRPEEECVGEGLACYP-CAHGHCWGPGPTQCVNCSQFLRGQEC  539

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC  600
              |||||||||||||||  |:||||| |||||||||||| ||||||||||||||||||||||
Db        540 VEECRVLQGLPREYVKDRYCLPCHSECQPQNGSVTCFGSEADQCVACAHYKDPPFCVARC  599

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG  660
              |||||||||:|||||| |||| |||||||||||| |||:|||||||||||:|||::||||
Db        600 PSGVKPDLSFMPIWKFADEEGTCQPCPINCTHSCADLDEKGCPAEQRASPVTSIIAAVVG  659

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN  720
              ||| ||:|:|  |||||||:||||||||||||||||||||||||||||||||||||||||
Db        660 ILLAVVVGLVLGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE-  718

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV  779
                    ||  ||  |    |||||||||||||||||||||||||||||||||||||||||
Db        719 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV  772

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM  839
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||:||||||
Db        773 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGM  832

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES  899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        833 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES  892

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM  959
              |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        893 IPPRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM  952

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        953 IMVKCWMIDSECRPRFRELVAEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1012

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              |||||||||||||||||||:| |||||  | ||||||||:|||:|||||||||||| |:|
Db       1013 MGDLVDAEEYLVPQQGFFCPEPTPGAGGTAHRRHRSSSTRNGGGELTLGLEPSEEEPPKS 1072

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              |||||||||||||||||||||||||||||||||: |||||||||||||| |||| ||||||
Db       1073 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPSQDPSPLQRYSEDPTVPLPPETDGKVAPLTC 1132

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLER-----AKTLSPGKNGVVKDVFAFG 1194
              |||||||||:| ||||   ||||| :|||||||||      |||||||||||||||||||
Db       1133 SPQPEYVNQPEVWPQPPLALEGPLPPSRPAGATLERPKTLSPKTLSPGKNGVVKDVFAFG 1192

Qy       1195 GAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254
              |||||||| |:| ||||||||||||||||||||||| |||:|||||:||||||||||
Db       1193 SAVENPEYLAPRGRAAPQPHPPPAFSPAFDNLYYWDQDPSERGSPPSTFEGTPTAENPEY 1252

Qy       1255 LGLDVPV 1261
              |||||||
Db       1253 LGLDVPV 1259


RESULT 2
Q9QX70
ID   Q9QX70      PRELIMINARY;      PRT;  1209 AA.
AC   Q9QX70;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor.
GN   EGFR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RX   MEDLINE=90258888; PubMed=2342466;
RA   Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA   Earp H.S.;
RT   "A truncated, secreted form of the epidermal growth factor receptor is
RT   encoded by an alternatively spliced transcript in normal rat tissue.";
RL   Mol. Cell. Biol. 10:2973-2982(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Petch L.A.;
RL   Submitted (NOV-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Guttridge K., Dawson T.L., Earp H.S.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M37394; AAF14008.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1209 AA;  134891 MW;  96FEE7F6CC1B7773 CRC64;

  Query Match             44.7%;  Score 3063;  DB 11;  Length 1209;
  Best Local Similarity   49.1%;  Pred. No. 1.5e-223;
```

```
  Matches' 631;  Conservative  169;  Mismatches  362;  Indels  122;  Gaps   29;

Qy          3 LAALCRWGLLLALLPPGA-ASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE  61
              ||||| |         ||    :|| ||  :|    : | |   |: ::  |:|| ||||
Db         15 LAALCAAG--------GALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMFNNCEVVLGNLE  66

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  ||||: |||| |||||| | | ::||: |:|:|| |:|: ||||||  |
Db         67 ITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAVLSN-- 124

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
                    | :  ||||| :|:| ||| | |    || ||  :|| |:||  ::  |:
Db        125 --------YGTNKTGLRELPMRNLQEILIGAVRFSNNPILCNMETIQWRDIV-QDVFLSN 175

Qy        182 TLIDTNRS-RACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              :|  |     |  | | |     |||    |:|| ||: :||  |: ||:|  |:|||| |
Db        176 MSMDVQRHLTGCPKCDPSCPNGSCWGRGEENCQKLTKIICAQQCSRRCRGRSPSDCCHNQ 235

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:||  |
Db        236 CAAGCTGPRESDCLVCHRFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKC 295

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              | ||: || |||    |    ||  |||  :|:||  || :|| |:|:    | ::  |
Db        296 PRNYVVTDHGSCVRACGPDYYEV-EEDGVSKCKKCDGPCRKVCNGIGIGEFK-DTLSINA 353

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | ::  |  | | |  || :| ||  : | || | :|:: :|::||||:| | |||::
Db        354 TNIKHFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPEN 413

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
              ||  |:||::|||   :| :|| : || |: ||||||:|:  |  :|  | :||: :
Db        414 WTDLHAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYAN 473

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: | :||  |:|      || | :|      |: ||:   |||| || ||:|    ||:
Db        474 TINWKKLFGTPNQKTKIMNNRAEKDCKATNHVCNPLCSSEGCWGPEPTDCVSCQNVSRGR 533

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||::| :|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||
Db        534 ECVDKCNILEGEPREFVENSECIQCHPECLPQTMNITCTGRGPDNCIKCAHYVDGPHCVK 593

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP-LTSIVSA 657
              ||||:  : : : :||| |    |  |  |||: |     |||  :|   | : || :
Db        594 TCPSGIMGENNTL-VWKFADANNVCHLCHANCTYGCAGPGLKGC--QQPEGPKIPSIATG 650

Qy        658 VVGILLVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
              :|| || :|: |  || |   ||:| :|| |:|||| |||||||| |||| :||||
Db        651 IVGGLLFIVV-VALGIGLFMRRRQLVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILK 709

Qy        717 EFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
              |      : : ::  : :    |||||:|||:|| |||||||| ||| ||||||||||||
Db        710 E------TEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDE 763

Qy        777 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 836
              ||||| | :|:| |||||||||||||:||||||||||:|||::  :||| |||||:|||
Db        764 AYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIA 823

Qy        837 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 896
              |||:|||| ||||||||||||||:| ||||||||||:||  :| ||||:||||||||||
Db        824 KGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMA 883

Qy        897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
              ||||| | :|||||||||||||||||||:|||||||| ||  :|||||||||||||||
Db        884 LESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTID 943

Qy        957 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLL 1015
              ||||||||||:: ||:|||: |||:|||||||::||| :| :   || || |||:|:
Db        944 VYMIMVKCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALM 1003

Qy       1016 EDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEE 1075
              |::|| |:|||:|||:||||||                 | ||
Db       1004 EEEDMEDVVDADEYLIPQQGFF----------------NSPST---------------- 1030

Qy       1076 APRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGY 1133
                |:|| |  | |:       :       |    : : ||||| |||  |  :   | :
Db       1031 -SRTPLLSSLSANSN----SSTVACINRNGSCRVKEDAFLQRYSSDPTSVLTEDNIDDTF 1085

Qy       1134 VAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAF 1193
              :       | |||:||  |  :|    : |:   :|           :||::   ::
Db       1086 L------PVPEYINQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PH 1128

Qy       1194 GGAVENPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PER 1236
                || ||||| | |       |  |  ||:   | |           :|       |:
Db       1129 SNAVSNPEYLNTAQ---------PTCLSSGFDSSALWIQKGSHQMSLDNPDYQQDFFPKE 1179

Qy       1237 GAPPSTFKGTPTAENPEYLGLDVP 1260
                |  ||| ||||| ||| :  |
Db       1180 AKPNGIFKG-PTAENAEYLRVAPP 1202
```

RESULT 3
Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maihle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maihle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.

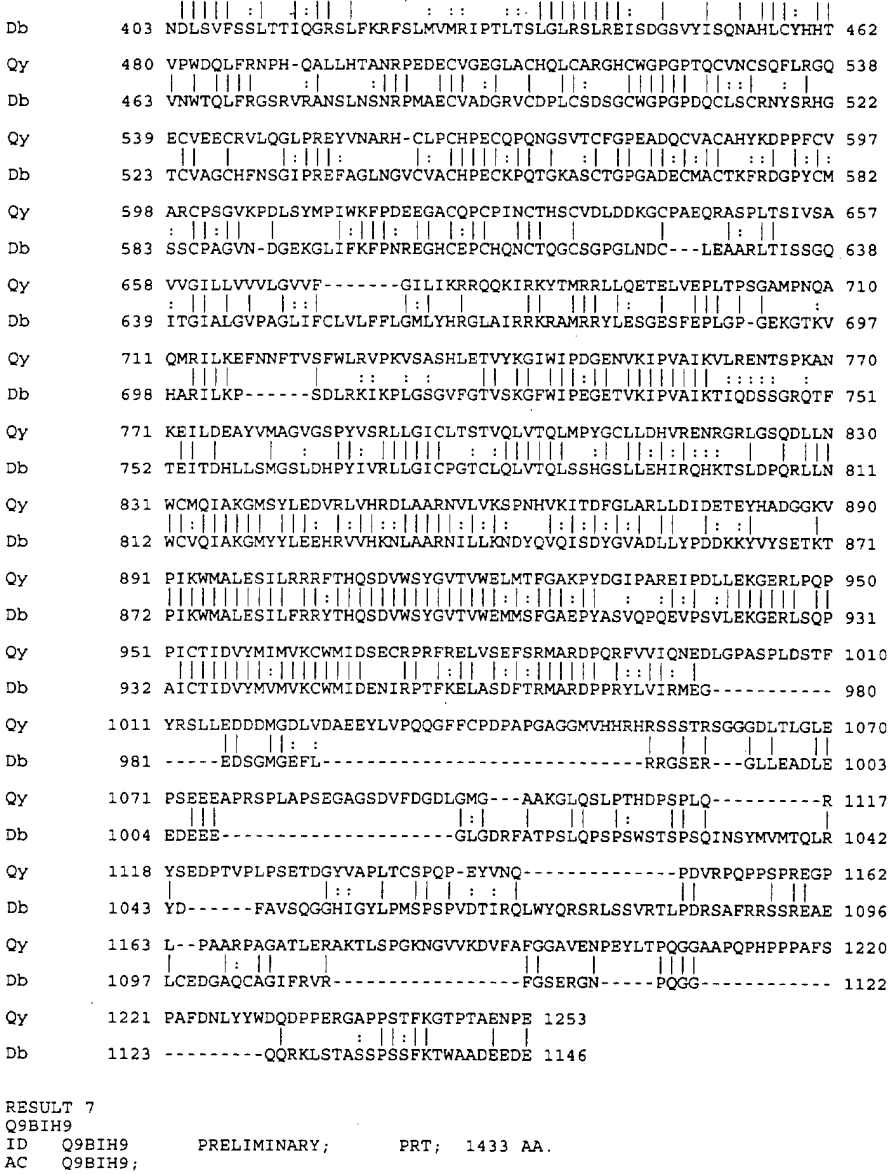
```
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase.
SQ   SEQUENCE   1210 AA;  134840 MW;  62CD021C9DE32E18 CRC64;

  Query Match             44.3%;  Score 3034;  DB 11;  Length 1210;
  Best Local Similarity   48.6%;  Pred. No. 2.4e-221;
  Matches  622;  Conservative  171;  Mismatches  368;  Indels  118;  Gaps   26;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 LVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNNF 721
Db        657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKE---- 711

Qy        722 TVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 781
Db        712 --TEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMA 769

Qy        782 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 841
Db        770 SVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNY 829

Qy        842 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 901
Db        830 LEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESIL 889

Qy        902 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 961
Db        890 HRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIM 949

Qy        962 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDM 1020
Db        950 VKCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDM 1009

Qy       1021 GDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSP 1080
Db       1010 EDVVDADEYLTPQQGFF----------------NSPST------------------SRTP 1035

Qy       1081 LAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLT 1138
Db       1036 LLSSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL---- 1087

Qy       1139 CSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVE 1198
Db       1088 --PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVG 1134

Qy       1199 NPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPS 1241
Db       1135 NPEYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNG 1185

Qy       1242 TFKGTPTAENPEYLGLDVP 1260
Db       1186 IFKG-PTAENAEYLRVAPP 1203
```

RESULT 4

```
Q9YH40
ID   Q9YH40      PRELIMINARY;      PRT;  1165 AA.
AC   Q9YH40;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase proto-oncogene.
GN   XMRK.
OS   Xiphophorus xiphidium.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8086;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RX   MEDLINE=98241172; PubMed=9582016;
RA   Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA   Altschmied J., Schartl M.;
RT   "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT   overexpression and mutational alterations.";
RL   Oncogene 16:1681-1690(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53471; AAD10500.2; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
```

```
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1165 AA;  129614 MW;  7F7EE38D8771A74E CRC64;

  Query Match             38.7%;  Score 2650;  DB 13;  Length 1165;
  Best Local Similarity   44.7%;  Pred. No. 3.3e-192;
  Matches  574;  Conservative  163;  Mismatches  389;  Indels  158;  Gaps   33;

Qy          1 MELAALCRWGLLLALLPPG-AAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQV 55
              :||  |    ||| ||  |   ||    :|| ||  :: :    : |   :: :| || |
Db          4 LELLEL----LLLLLLSIGRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNV 56

Qy         56 VQGNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALA 115
              |  |||:||    |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: |
Db         57 VLENLEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLL 116

Qy        116 VLDNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHK 175
              |: |    | ::|     ||::||| :||||| ||| :  || ||  :|| | ||  |
Db        117 VMSNYQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDK 173

Qy        176 NNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTD 234
              :   : ||     | |  | | |    ||      ||  |: :||  |  ||:|| | |
Db        174 TSNPTMNLIPHAFERQCQKCDPGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPID 233

Qy        235 CCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
              ||:| || ||||||: :||||  ||  | |:  ||     |:  : : :  ||  :|||||:
Db        234 CCNEHCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAA 293

Qy        295 CVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGM----QYI 350
              ||  || ||: |: |:|    |      ||  |:| : |: |    | :|| |:|:      |
Db        294 CVKECPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSLSNTI 351

Qy        351 KANSKFIGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
                ||  ||     |: | || | :     ||:|||      |: || |    |::||||||
Db        352 AVNSTNIG----SFSNCTKINGDIILNRNSFEGDPHYKIGPMDPEHLWNLTTVKEITGYL 407

Qy        411 YISAWPDSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIH 469
              |  ||:::  |||||||::||||   :  :| : :|   : |||| ||:|: :|  ::
Db        408 VIMWWPENMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVSHLQWLGLSSLKEVSAGNVILK 467

Qy        470 HNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCV 529
              :   | :  |: | :|||:  |:: : |          |    |:  |:   |||||| ||
Db        468 NTPQLRYASTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCV 520

Qy        530 NCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAH 589
              :    ||  ||  | :||| |||       |: || ||   |  |:|:||    | |||
Db        521 SCLHVDRGGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKCAH 580

Qy        590 YKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS 649
              ::| | |: ||| |:  |   : |||: |: | |||  |||  |      ||  :   |
Db        581 FQDGPQCIPRCPHGMLGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGD-IVS 638

Qy        650 PLTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQ 709
              :  |  | |:|: |:: ::  :|::||: | || |:|||||| |||||||||  |||
Db        639 HSSLAVGLVSGLLITVIVALLIVVLLRRRRIK-RKRTIRRLLQEKELVEPLTPSGQAPNQ 697

Qy        710 AQMRILK--EFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPVAIKVLRENTSP 767
              | :|||  ||    |       : :    |||||:| |||||::|||||||||| |||
Db        698 AFLRILKETEFKKDRV--------LGSGAFGTVYKGLWNPDGENIRIPVAIKVLREATSP 749

Qy        768 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 827
              | |:|:|||||||| |  |:| |||||||| |||||||||||||||:||::: |:  |
Db        750 KVNQEVLDEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQW 809

Qy        828 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 887
              |||||:|||||:|||:  |||||||||||||:|:||||||||||::||  || || | |
Db        810 LLNWCVQIAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQAHG 869

Qy        888 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 947
              |||||||||||||:  :|||||||||||||||||||:||||||||:||  :|| ||||
Db        870 GKVPIKWMALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERL 929

Qy        948 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLD 1007
              ||||||||:||||::||||||   |||||||| |||:|||||  |::|||        || |
Db        930 PQPPICTIEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSPSD 986

Qy       1008 STFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTL 1067
                :  ||  ||  |:|||:|||:                                |::  : | |
Db        987 RRLFSRLLSSDD--DVVDADEYLL-------------------RYKRIN-RQGS----- 1018

Qy       1068 GLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PL 1126
                          |  |  |                  |  : |  ||  |||  |
Db       1019 ----------EPCIPPNGH-------------------PVRENSIALRYISDPTQNAL 1047

Qy       1127 PSETDGYVAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATL 1173
                : ||:          ||||||               |:  :     ||:  ::: |
Db       1048 EKDLDGH----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNF 1097

Qy       1174 ERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDP 1233
                | : |:  :| :         |::::|:|   | |           :  ||
Db       1098 SRPEYLNTNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------L 1132

Qy       1234 PERGAPPSTFKGTPTAENPEYLGL 1257
                |: ||     | ||| |||||
Db       1133 PQTGALTGNGMFLPAAENLEYLGL 1156


RESULT 5
Q9W6F6
ID   Q9W6F6      PRELIMINARY;      PRT;  1137 AA.
AC   Q9W6F6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase (Fragment).
GN   ERBB4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HINDBRAIN;
RX   MEDLINE=99263203; PubMed=10328884;
RA   Dixon M., Lumsden A.;
RT   "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
RT   embryonic chick hindbrain.";
RL   Mol. Cell. Neurosci. 13:237-258(1999).
DR   EMBL; AF121963; AAD31764.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001368; TNFR_c6.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 1.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW   Kinase; Tyrosine-protein kinase.
FT   NON_TER       1      1
SQ   SEQUENCE   1137 AA;  127927 MW;  4D616436F87DC84F CRC64;

  Query Match             38.1%;  Score 2612.5;  DB 13;  Length 1137;
```

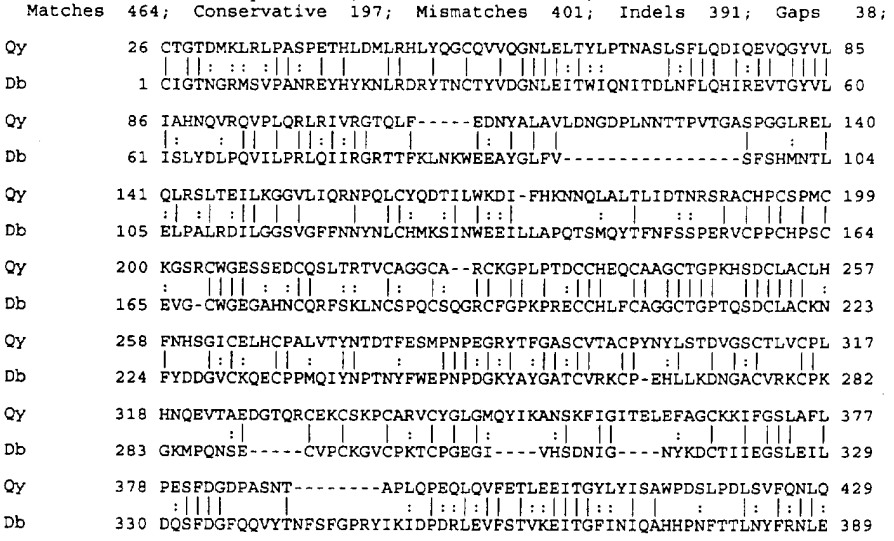
```
  Best Local Similarity   45.6%;  Pred. No. 2.2e-189;
  Matches  523;  Conservative  175;  Mismatches  363;  Indels   85;  Gaps   27;

Qy        161 LCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVC 220
              ||: ||| |:||          ||: || |  |  |  | | |||| :   ||:||:|||
Db          3 LCFADTIHWQDIVRNPWASNFTLVPTNGSSGCGRCHKSCTG-RCWGPTENHCQTLTKTVC 61

Qy        221 AGGC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTF 279
              |  |  || ||  :|||| :|| ||:||| :|| ||::|| || |   ||    ||  ||
Db         62 AEQCDGRCYGPYVSDCCHRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTF 121

Qy        280 ESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCA 339
              :   |   :||:|| ||  ||:|:: |  ||   ||    ||  |:| : |: |:  |
Db        122 QLEHNHNAKYTYGAFCVKKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICP 179

Qy        340 RVCYGLGMQYIKANSKFIGITELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQ 398
              : | |:|   :  ::: :  : :: |  | || |:| ||     |||     | : ||:|
Db        180 KACDGIGTGSL-VSAQTVDSSNIDKFINCTKINGNLIFLVTGIHGDPYHTIAAINPEKLN 238

Qy        399 VFETLEEITGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSL 458
              :|:|: |||||| | :||::: |  || ||  | || |::|   | |:  ||: |  :||
Db        239 IFQTVREITGYLNIQSWPENMTDFRVFSNLVTIGGRALYSGLSLLILKQQGITSLQFQSL 298

Qy        459 RELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARG 518
              ::: :|   |  |::||: ||| |  ||  | |   :   |:  : |  :|: |::||:
Db        299 KQISAGNIYITDNSNLCYYHTVNWTSLFSTPSQKTVIHRNKKAENCTADGMVCNELCSSD 358

Qy        519 HCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCF 577
              |||||| ||::| :|:||: |:|  | :  |  ||: |   |: | |:|:  ::  :||:
Db        359 GCWGPGPDQCLSCKRFIRGRTCIESCNLYDGEFREFANGSVCMECDPQCEKMEDNMITCY 418

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |  | |:|| | || :|| |::   |:  |:|: ||:  | ||  |||  |
Db        419 GPGPDHCTKCFHFKDGPNCVEKCPDGLQGANSF--IFKYADEDRECHPCHPNCTQGCRGP 476

Qy        638 DDKGC-----------PAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKY 685
                  |           |   | :||  | : |: |: ::|::|: | : ::|:  | :|
Db        477 ASHDCIYYPWTRQSTLPQHAR-TPL--IAAGVIGGLFIIVIMGLTFAVYVRRKSIK-KKR 532

Qy        686 TMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIW 745
              :|| | |||||||||||  |||||:|||||      :   ||  : :    |||||||
Db        533 ALRRFL-ETELVEPLTPSGTAPNQAQLRILKE------TELKRVKVLGSGAFGTVYKGIW 585

Qy        746 IPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQ 805
              :|:|| |||||||:| | | |||| | :||| :||  :   |:: |||:||: |:|||||
Db        586 VPEGETVKIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQ 645

Qy        806 LMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHV 865
              |||:|||||:| |::  :||| |||||:|||||| ||: ||||||||||||||||||||
Db        646 LMPHGCLLDYVHEHKDNIGSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHV 705

Qy        866 KITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFG 925
              |||||||||||: || ||:|||||:|||||||| |  |:|||||||||||||:|||||||
Db        706 KITDFGLARLLEGDEKEYNADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFG 765

Qy        926 AKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRM 985
              ||||||| |||||||||||||||||||||||||:|||||||:: ||:|:|| :|||||
Db        766 GKPYDGIPTREIPDLLEKGERLPQPPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRM 825

Qy        986 ARDPQRFVVIQNED-LGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPG 1044
              ||||||::||| :| :   || ||  |:::||:::|: |::|||||||| | |  | |
Db        826 ARDPQRYLVIQGDDRMKLPSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYT 884

Qy       1045 AGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS--PLAP-SEGAGSDVFDGDLGMGAA 1101
              :   :     | ||     |: |    ||    | || ::|| :::|:     |
Db        885 SRTRIDSNRNQFVYRDGGYAAEQGV-PMPYRAPGCIIPEAPVAQGATAEIFEDTCCNGTL 943

Qy       1102 KGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQ 1154
              :   :     : |  |||| |||| :|        : |||: |:   |: :|:|  :  |
Db        944 RKQVATLAKEDSSTQRYSADPTVFIPERVIRGELDEDGYMTPMRDKPKTDYLNPVEENPF 1003

Qy       1155 PPSPREGPLPAA-RPAGATLERAKTLSPGKNGVVKDVF--------AFGGAVENPEYLTP 1205
                 : | | |   |                :|      :| :         |   :|| |||

Db       1004 VSRRKNGDLQAVDNPEYHN-------APNGQPKAEDEYVNEPLYLNTFANTLENAEYL-- 1054

Qy       1206 QGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGA--PPSTFKGTPT-------------A 1249
                    : : |     ||||  ||:  | |     |   :  |              |
Db       1055 ------KNNLPEKAKKAFDNPDYWNHSLPPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVA 1108

Qy       1250 ENPEYL 1255
                |||||
Db       1109 ENPEYL 1114


RESULT 6
P79754
ID   P79754      PRELIMINARY;      PRT;  1328 AA.
AC   P79754;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ErbB3.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99177347; PubMed=10077531;
RA   Gellner K., Brenner S.;
RT   "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT   rubripes.";
RL   Genome Res. 9:251-258(1999).
DR   EMBL; AF056116; AAC34391.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   1328 AA;  148613 MW;  A333039258B647E9 CRC64;

  Query Match             32.5%;  Score 2226;  DB 13;  Length 1328;
  Best Local Similarity   39.4%;  Pred. No. 6.5e-160;
  Matches  510;  Conservative  158;  Mismatches  427;  Indels  198;  Gaps   33;

Qy          9 WGLLLALLPP--GAASTQ----VCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLEL 62
              | |:|  :     |||:|    || ||  |   | |   :: :  |:||::: ||||:
Db          4 WRLILMCVASRLRAASSQTQEAVCPGTQNGLSSTGSQENQYNLNKDRYKGCEIIMGNLEI 63

Qy         63 TYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDP 122
              | : :|   |||: |:|| |||||| |  :::|| :||::||  |:|  :||:|  |
Db         64 TQIESNWDFSFLKTIREVTGYVLIAMNHFQEIPLGQLRVIRGNSLYERRFALSVFLN--- 120

Qy        123 LNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALT 182
                  |  |  | || :| | :|||| ||| |  |  | |  :|:||  :||   :
Db        121 ----YPKDG--PSGLNQLGLMNLTEILDGGVQIINNKYLRYGPWVYWRDII-RNNDAPIE 173

Qy        183 LIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCA 241
              :          ||      | |: |||   : || ||:||||  |  || |  | |||| :||
Db        174 IQFNGERGVCH---KSC-GNYCWGPGKDQCQILTKTVCAPQCNDRCFGTSPRDCCHIECA 229

Qy        242 AGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY 301
              ||| ||  :|| ||  || || |  ||  : ||  ||:   ||  :| :|: ||: ||
Db        230 AGCKGPLDTDCFACRLFNDSGACVPQCPQTLIYNKQTFQMETNPNAKYQYGSICVSQCPT 289
```

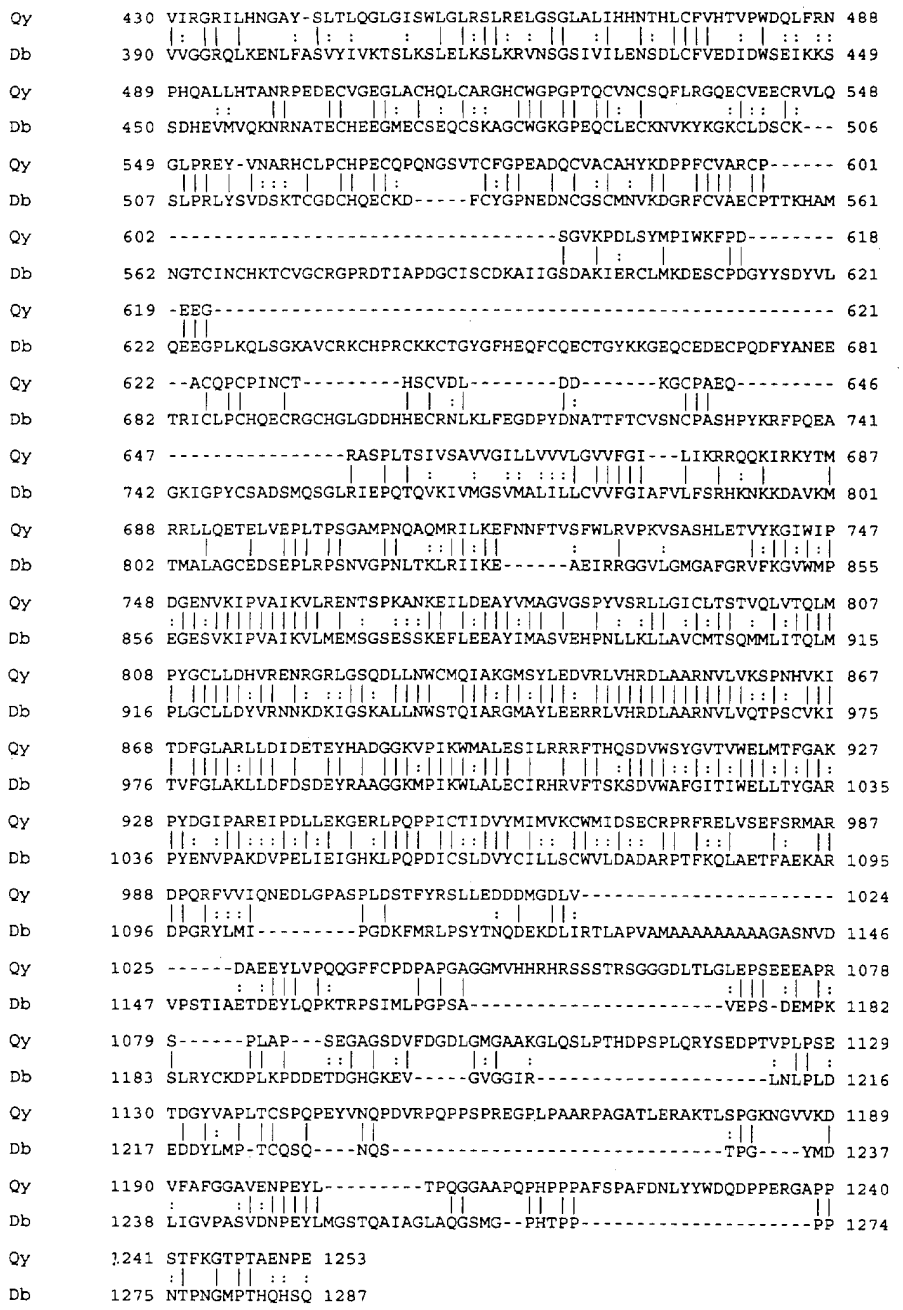


RESULT 7
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 28.6%; Score 1959.5; DB 5; Length 1433;
Best Local Similarity 31.9%; Pred. No. 1.3e-139;
Matches 464; Conservative 197; Mismatches 401; Indels 391; Gaps 38;





RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.3%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 1.1e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027080; AAH27080.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   367 AA;  40163 MW;  0BE03395F9E101B0 CRC64;

  Query Match           25.4%;  Score 1739;  DB 11;  Length 367;
  Best Local Similarity   88.0%;  Pred. No. 9.5e-124;
  Matches  323;  Conservative   11;  Mismatches   33;  Indels    0;  Gaps    0;

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||:||:||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSL 120

Qy       1015 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1074
              |||||||:|||||||||||||||| |||| | |    | |||||| |||||:||||||||||
Db        121 LEDDDMGELVDAEEYLVPQQGFFSPDPALGTGSTAHRRHRSSSARSGGGELTLGLEPSEE 180

Qy       1075 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1134
              | |||||||||||||||||||| :| |||||  || ||||||||||:||| ||||||
Db        181 EPPRSPLAPSEGAGSDVFDGDLAVGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYV 240

Qy       1135 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1194
              ||| ||||||||||:|||| |   ||| | |||||||| ||||||||||||||||||||
Db        241 APLACSPQPEYVNQPEVRPQSPLTPEGPPPPIRPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1195 GAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1254
              |||||||| |: | | |||| ||||||||||||||:  |:| |||||:||||||||||
Db        301 GAVENPEYLAPRAGTASQPHPSPAFSPAFDNLYYWDQNSSEQGPPPSTFEGTPTAENPEY 360

Qy       1255 LGLDVPV 1261
              |||||||
Db        361 LGLDVPV 367


RESULT 10
Q8WYV0
ID   Q8WYV0      PRELIMINARY;      PRT;   412 AA.
AC   Q8WYV0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 44.7 kDa protein.
GN   PP3659.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA   Wan D.F., Gu J.R.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF318349; AAL55856.1; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   412 AA;  44702 MW;  034397FF3F27D2BC CRC64;

  Query Match           24.8%;  Score 1697.5;  DB 4;  Length 412;
  Best Local Similarity   80.5%;  Pred. No. 1.6e-120;
  Matches  330;  Conservative    5;  Mismatches   22;  Indels   53;  Gaps    4;

Qy        895 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 60

Qy        955 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 120

Qy       1015 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 180

Qy       1075 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYV 240

Qy       1135 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFG 1194
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        241 APLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFG 300

Qy       1195 GAVENPEYLTPQGGAAPQP------------------------HPPPA---FSPAFDNL 1226
              |||||||||||||||  |                        || |:        :
Db        301 GAVENPEYLTPQGGAALSPTLLLPSAQPSTTSITGTRTHQSGGLHPAPSKGHLRQRTQST 360

Qy       1227 YYWD-QDPPER------------------------GAPPSTFKGTPTAEN 1251
              : |  |  ||                        ||:| :||  | |
Db        361 WVWTCQCEPEGQVRRSPDVSSGSREGLTSAGIKRWEGPPTTSRGTCHARN 410


RESULT 11
Q86712
ID   Q86712      PRELIMINARY;      PRT;   729 AA.
AC   Q86712;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyprotein.
GN   POLYPROTEIN.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60725.1; -.
DR   HSSP; P03322; 1A6S.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR004028; Retro_M.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02813; Retro_M; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   729 AA;  80649 MW;  84D2F6914EFE1D63 CRC64;

  Query Match           23.9%;  Score 1639;  DB 15;  Length 729;
  Best Local Similarity   52.5%;  Pred. No. 1e-115;
```

```
  Matches  346;  Conservative   79;  Mismatches  126;  Indels  108;  Gaps   16;

Qy        569 PQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPI 628
              |:  :    ||  | |: |||: | | ||  ||:||  :   : :||: |    || |
Db        141 PEETATPKTGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHP 197

Qy        629 NCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTM 687
              |||  |     :|||      |   || : || |:| :||:|:  |: ::||   :|| |:
Db        198 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTL 253

Qy        688 RRLLQETELVEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIP 747
              |||||| |||||||||  |||| :|||||      : : :|  : :    |||||:|||
Db        254 RRLLQERELVEPLTPSGEAPNQAHLRILKE------TEFKKVKVLGSGAFGTVYKGLWIP 307

Qy        748 DGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLM 807
              :|| |||||||| ||| |||||||||||||||| | :| | ||||||||||||||:||||
Db        308 EGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLM 367

Qy        808 PYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKI 867
              |||||||::||::  :||| ||||:||||||:|||: |||||||||||||:| ||||
Db        368 PYGCLLDYIREHKDNIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKI 427

Qy        868 TDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAK 927
              ||||||:||  || ||||:|||||||||||||| | :|||||||||||||||||||:|
Db        428 TDFGLAKLLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSK 487

Qy        928 PYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMAR 987
              ||||||| ||  :|||||||||||||||||||||||:: ||:|||::|||:|||
Db        488 PYDGIPASEISSVLEKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMAR 547

Qy        988 DPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAG 1046
              || |::||| :| :   || || |||:|:|::|| |:|||:|||| ||||
Db        548 DPPRYLVIQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF--------- 598

Qy       1047 GMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAA 1101
                | ||                          |:||     | |  : ::  |
Db        599 -------NSPST-----------------SRTPLLSSLSATSNNSATNCID-------- 625

Qy       1102 KGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPR 1159
              :  |  |  : | :|||| |||     |:  ||::      | ||||||  : |: ||
Db        626 RNGQGHPVREDSFVQRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS-- 675

Qy       1160 EGPLPAARPAGATLERAKTLSPGKNGVVKDVF--------------AFGGAVENPEYL 1203
                              :  :| :  ::                 :   ||:|||||
Db        676 -------------------TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 715


RESULT 12
Q86714
ID   Q86714      PRELIMINARY;      PRT;   567 AA.
AC   Q86714;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   V-erbB protein (Fragment).
GN   V-ERBB.
OS   Avian rous-associated virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94203659; PubMed=8152791;
RA   Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA   Johnsson A., Beug H.;
RT   "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT   evolution of distinct viral genomes carrying mutant v-erbB genes with
RT   different transforming capacities.";
RL   Oncogene 9:1307-1320(1994).
DR   EMBL; S69372; AAC60727.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Tyrosine-protein kinase.
FT   NON_TER       1      1
SQ   SEQUENCE   567 AA;  63390 MW;  C6D9CBA7ADF725E1 CRC64;

  Query Match             23.9%;  Score 1637;  DB 15;  Length 567;
  Best Local Similarity   53.1%;  Pred. No. 1e-115;
  Matches  345;  Conservative   77;  Mismatches  120;  Indels  108;  Gaps   16;

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |: |||: | | ||  ||:||  :   : :||: |    || |  |||  |
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL 696
                :|||     |    || : || |:| :||:|:  |: ::||   :|| |:|||||| ||
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQEREL 113

Qy        697 VEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPV 756
              ||||||||  |||| :|||||      : : :|  : :    |||||:|||:|| |||||
Db        114 VEPLTPSGEAPNQAHLRILKE------TEFKKVKVLGSGAFGTVYKGLWIPEGEKVKIPV 167

Qy        757 AIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHV 816
              ||| ||| ||||||||||||||| | :| | ||||||||||||||:|||||||||::
Db        168 AIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCLLDYI 227

Qy        817 RENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLL 876
              ||::  :||| |||||:||||||:|||: |||||||||||||||:| ||||||||||:||
Db        228 REHKDNIGSQYLLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLL 287

Qy        877 DIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPARE 936
              || ||||:|||||||||||||| | :|||||||||||||||||||:|||||||| |
Db        288 GADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASE 347

Qy        937 IPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ 996
              |  :|||||||||||||||||||||||||:: ||:|||::|||:|||| |::|||
Db        348 ISSVLEKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQ 407

Qy        997 -NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRS 1055
               :| :   || || |||:|:|::|| |:|||:|||| ||||                |
Db        408 GDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NS 451

Qy       1056 SSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTH 1110
                ||                |:||     | |  : ::  |          :  | |
Db        452 PST-----------------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVR 485

Qy       1111 DPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARP 1168
              : | :|||| |||     |:  ||::      | ||||||  : |: ||
Db        486 EDSFVQRYSSDPTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS----------- 526

Qy       1169 AGATLERAKTLSPGKNGVVKDVF--------------AFGGAVENPEYL 1203
                          :  :| :  ::                 :   ||:|||||
Db        527 ----------TAMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 566


RESULT 13
Q64895
ID   Q64895      PRELIMINARY;      PRT;   962 AA.
AC   Q64895;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Gag,v-erb-A,v-erb-B protein.
GN   GAG,V-ERB-A,V-ERB-B.
OS   Avian erythroblastosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=90206603; PubMed=1969616;
RA   Bruskin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT   "Six amino acids from the retroviral gene gag greatly enhance the
RT   transforming potential of the oncogene v-erb-B.";
RL   Oncogene 5:15-24(1990).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; X52209; CAA36459.1; -.
DR   EMBL; X52211; CAA36459.1; JOINED.
DR   HSSP; P10828; 2NLL.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW   Transcription regulation; Transferase; Tyrosine-protein kinase;
KW   Zinc-finger.
SQ   SEQUENCE   962 AA;  108320 MW;  3C5AED791E4E95CE CRC64;

  Query Match           23.0%;  Score 1575.5;  DB 15;  Length 962;
  Best Local Similarity  49.7%;  Pred. No. 1.1e-110;
  Matches  347;  Conservative   76;  Mismatches  150;  Indels  125;  Gaps   19;

Qy        541 VEECRVLQGLPRE-YVNAR-HCLP----------------------------CHPECQ 568
Db        354 IEKCQESYLLAFEHYINYRKHNIPHFWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413

Qy        569 PQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPI 628
Db        414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHP 465

Qy        629 NCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTM 687
Db        466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTL 521

Qy        688 RRLLQETELVEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIP 747
Db        522 RRLLQERELVEPLTPSGEAPNQAHLRILKE------TEFKKVKVLGFGAFGTVYKGLWIP 575

Qy        748 DGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLM 807
Db        576 EGEKVTIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLM 635

Qy        808 PYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKI 867
Db        636 PYGCLLDYIREHKDNIGSQYLLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKI 695

Qy        868 TDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAK 927
Db        696 TDFGLAKQLGADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSK 755

Qy        928 PYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMAR 987
Db        756 PYDGIPASEISSVLEKGERLPQPPICTIDVYMIMVKCWMSGADSRPKFRELIAEFSKMAR 815

Qy        988 DPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAG 1046
Db        816 DPPRYLVIQGDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF--------- 866
```

```
Qy       1047 GMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQS 1106
Db        867 -------NSPST-----------------SRTPLLSSLSATSN-------NSATKCIDR 894

Qy       1107 LPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAA 1166
Db        895 NGGH----------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIY 930

Qy       1167 RPAGAT-LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203
Db        931 NYISLTAISKLPMDSRYQN-------SHSTAVDNPEYL 961


RESULT 14
Q85468
ID   Q85468      PRELIMINARY;      PRT;   545 AA.
AC   Q85468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS   Avian erythroblastosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11861;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88217326; PubMed=2897102;
RA   Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT   "Common site of mutation in the erbB gene of avian erythroblastosis
RT   virus mutants that are temperature sensitive for transformation.";
RL   Oncogene Res. 1:265-278(1987).
DR   EMBL; X06943; CAA30024.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   545 AA;  60899 MW;  140DCE8CCA0F8AF4 CRC64;

  Query Match           22.9%;  Score 1567;  DB 15;  Length 545;
  Best Local Similarity  52.7%;  Pred. No. 2e-110;
  Matches  334;  Conservative   73;  Mismatches  129;  Indels   98;  Gaps   16;

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL 696
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQEREL 113

Qy        697 VEPLTPSGAMPNQAQMRILKEFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENVKIPV 756
Db        114 VEPLTPSGEAPNQAHLRILKE------TEFKKVKVLGFGAFGTVYKGLWIPEGEKVTIPV 167

Qy        757 AIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHV 816
Db        168 AIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYI 227

Qy        817 RENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLL 876
Db        228 REHKDNIGSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQL 287

Qy        877 DIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPARE 936
Db        288 GADEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASE 347
```

```
Qy        937 IPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ 996
Db        348 ISSVLEKGERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQ 407

Qy        997 -NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRS 1055
Db        408 GDERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NS 451

Qy       1056 SSTRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTH 1110
Db        452 PST-----------------SRTPLLSSLSATSNNSATNCIDRNGG------------H 481

Qy       1111 DPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAG 1170
Db        482 ---------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQIQIYNYIS 517

Qy       1171 AT-LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1203
Db        518 LTAISKLPMDSRYQN-------SHSTAVDNPEYL 544
```

RESULT 15
Q9WVF5
ID Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maihle N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maihle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.0%; Score 1506.5; DB 11; Length 655;
Best Local Similarity 44.5%; Pred. No. 1e-105;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

```
Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC 642
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGC 636
```

Search completed: July 22, 2003, 09:24:40
Job time : 53.0485 secs

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 37.1104 Seconds
(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-695-709-14
Perfect score: 6853
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6647	97.0	1255	21	AAY92620	Human heregulin 2
2	6647	97.0	1255	22	AAE12130	Human tyrosine kin
3	6647	97.0	1255	22	AAB60167	HER2 transgene pla
4	6647	97.0	1255	23	AAU74545	Human HER2 (ErbB2)
5	6641	96.9	1255	17	AAW01111	HER-2/neu protein.
6	6641	96.9	1255	20	AAW92406	Human HER-2/neu on
7	6641	96.9	1255	21	AAB21198	Human HER-2/neu pr
8	6641	96.9	1255	21	AAY84780	Amino acid sequenc
9	6641	96.9	1255	22	AAB85458	Human HER-2/neu pr
10	6641	96.9	1255	22	AAG88267	HER2/neu amino aci
11	6641	96.9	1255	23	AAE24067	Human Her-2 protei
12	6641	96.9	1255	23	AAE20479	Human Her-2/neu pr
13	6641	96.9	1255	23	AAM51143	Human Her-2/neu on
14	6641	96.9	1255	23	AAU77114	Human Her-2/neu po
15	6598	96.3	1433	14	AAR39568	Sequence of c-erbB
16	6477	94.5	1223	23	AAU98923	Human breast cance
17	6324	92.3	1200	21	AAB21208	Human HER-2/neu pr
18	5854.5	85.4	1256	21	AAB21199	Rat HER-2/neu prot
19	5854.5	85.4	1256	23	AAM51144	Rat Her-2/neu onco
20	5829.5	85.1	1256	21	AAB21206	Mouse Her-2/neu pr
21	5829.5	85.1	1256	22	AAG62860	Amino acid sequenc
22	5829.5	85.1	1256	23	AAM51151	Mouse Her-2/neu on
23	4811	70.2	919	21	AAB21203	Human HER-2/neu fu
24	4811	70.2	919	23	AAM51148	Her-2/neu extracel
25	4061.5	59.3	920	23	AAM51152	Mouse Her-2/neu ex
26	4061.5	59.3	926	23	AAM51153	Mouse Her-2/neu ex
27	3695	53.9	712	21	AAB21204	Human HER-2/neu fu
28	3695	53.9	712	23	AAM51149	Her-2/neu extracel
29	3552	51.8	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	51.8	653	21	AAB21200	Extracellular HER-
31	3550	51.8	653	23	AAM51145	Human Her-2/neu on
32	3512	51.2	645	22	AAB60408	Human ErbB2 oncopr
33	3512	51.2	645	22	AAB61593	Human ErbB2 extrac
34	3447	50.3	951	21	AAY44993	DC8scFv-erbB2EC fu
35	3344	48.8	624	11	AAR08222	Extracellular port
36	3065	44.7	1210	21	AAB19259	Amino acid sequenc
37	3065	44.7	1210	21	AAY50616	Human EGF receptor
38	3065	44.7	1210	23	AAE23019	Human Her-1 protei
39	3065	44.7	1210	23	AAM50768	Human epidermal gr
40	3063	44.7	1210	22	AAB68420	Amino acid sequenc
41	3053.5	44.6	654	21	AAB21205	Rat HER-2/neu prot
42	3053.5	44.6	654	23	AAM51150	Rat Her-2/neu onco
43	3024	44.1	1210	23	ABP51768	Human epidermal gr
44	2997	43.7	583	23	AAE20483	Human protein for
45	2997	43.7	587	23	AAE20481	Human protein for

ALIGNMENTS

RESULT 1
AAY92620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region 5..25
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region 59..73
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region 103..117
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region 149..163
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Domain 174..323

```
FT                   /label= Cysteine_rich_domain
FT   Region          210..224
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          250..264
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          324..483
FT                   /label= Ligand_binding_domain
FT   Region          325..339
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          369..383
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          465..479
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          484..623
FT                   /label= Cysteine_rich_domain
FT   Region          579..593
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          624..654
FT                   /label= Transmembrane_domain
FT   Region          632..652
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          653..667
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          655..1010
FT                   /label= Tyrosine_kinase_domain
FT   Region          661..675
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          695..709
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Region          710..730
FT                   /label= insertion_region
FT                   /note= "suitable for foreign epitope insertion"
FT   Domain          1011..1235
FT                   /label= C-terminal_domain
XX
PN   WO200020027-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;   99WO-DK00525.
XX
PR   05-OCT-1998;   98DK-0001261.
PR   20-OCT-1998;   98US-0105011.
XX
PA   (MEBI-) M & E BIOTECH AS.
XX
PI   Steinaa L,  Mouritsen S,  Nielsen KG,  Haaning J,  Leach D,  Dalum I;
PI   Gautam A,  Birk P,  Karlsson G;
XX
DR   WPI; 2000-349917/30.
DR   N-PSDB; AAA09455.
XX
PT   Inducing immune responses to weakly immunogenic, tumor associated
PT   peptide antigens for the treatment of breast and prostate cancer
XX
PS   Claim 62; Page 193-198; 220pp; English.
XX
CC   This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC   Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC   response.  Subdominant CTL epitopes, antibody binding regions and
CC   cysteine residues involved in disulfide bonds are preserved in the
CC   immunogenized forms.  Regions suitable for the insertion of foreign T
CC   helper epitopes were identified (see features table).  The method
CC   is used for inducing immune responses against weakly immunogenic
CC   cell-associated peptide antigens (PA) such as those associated with
CC   cancers (self-proteins), e.g. human prostate specific membrane antigen
CC   (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC   The method comprises effecting simultaneous presentation by antigen
CC   producing cells (APCs) of the animals immune system of: (1) at least 1
CC   CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC   B-cell group derived from the cell-associated PA; and (2) at least 1
CC   first T helper cell group which is foreign to the animal.  Analogues of
CC   human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC   part of all known and predicted CTL and B-cell epitopes of the respective
CC   PA and including at least one foreign T helper epitope are also claimed.
CC   The method is used to treat prostate, prostate/breast or breast cancer
CC   when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ   Sequence   1255 AA;

  Query Match             97.0%;  Score 6647;  DB 21;  Length 1255;
  Best Local Similarity   97.5%;  Pred. No. 0;
  Matches 1230;  Conservative    4;  Mismatches   20;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||  |     ||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773
```

```
Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 2
AAE12130
ID   AAE12130 standard; Protein; 1255 AA.
XX
AC   AAE12130;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human tyrosine kinase-type receptor, HER-2.
XX
KW   Therapeutic compound; major histocompatibility complex; vaccine;
KW   antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW   adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW   antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          774..782
FT                   /note= "Antigenic epitope"
XX
PN   WO200168677-A2.
XX
PD   20-SEP-2001.
XX
PF   16-MAR-2001; 2001WO-US40328.
XX
PR   16-MAR-2000; 2000US-0527487.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Nicolette CA;
XX
DR   WPI; 2001-616284/71.
DR   N-PSDB; AAD19731.
XX
PT   Novel synthetic therapeutic compound for inducing immune response and
PT   for use in adoptive immunotherapy, has enhanced binding to major
PT   histocompatibility molecules and enhanced immunoregulatory properties
PT   -
XX
PS   Claim 4; Page 63-67; 69pp; English.
XX
CC   The invention relates to synthetic therapeutic compounds (antigenic
CC   peptides) with enhanced binding to major histocompatibility complex
CC   (MHC) molecules and enhanced immunoregulatory properties relative
CC   to their natural counterparts. Compounds of the invention are useful
CC   for inducing an immune response in a subject and for use in adoptive
CC   immunotherapy. They are useful as components of anti-cancer vaccines
CC   and to expand immune effector cells that are specific for cancers
CC   characterised by expression of the breast cancer antigen, HER-2.
CC   Polynucleotides that encode peptides of the invention are useful as
CC   hybridisation probes and as primers for the detection of genes of gene
CC   transcripts that are expressed in antigen presenting cells (APCs), to
CC   confirm transduction of polynucleotides into host cells. The present
CC   sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC   of the invention are designed based on the HER-2 antigenic peptide
CC   (774-782).
XX
SQ   Sequence   1255 AA;

  Query Match             97.0%;  Score 6647;  DB 22;  Length 1255;
  Best Local Similarity   97.5%;  Pred. No. 0;
  Matches 1230;  Conservative    4;  Mismatches   20;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
```

```
Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||  |    ||||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 3
AAB60167
ID   AAB60167 standard; Protein; 1255 AA.
XX
AC   AAB60167;
XX
DT   03-APR-2001  (first entry)
XX
DE   HER2 transgene plasmid construct encoded protein.
XX
KW   Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW   antibody.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200100244-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-US17229.
XX
PR   25-JUN-1999;   99US-0141316.
PR   16-MAR-2000; 2000US-0189844.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Erickson S,  Schwall R;
XX
DR   WPI; 2001-061962/07.
DR   N-PSDB; AAF24297.
XX
PT   Treating tumors, particularly breast cancers, which overexpress an ErbB
PT   receptor and does not respond to an anti-ErbB antibody, comprises
PT   conjugating the antibody to a maytansinoid -
XX
PS   Example 3; Fig 4; 92pp; English.
XX
CC   The present invention provides a method of treating cancer by
CC   administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC   particular, the antibody is directed against ErbB2 (also known as HER2
CC   and p185neu). The method is particularly useful in the treatment of
CC   breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC   colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match             97.0%;  Score 6647;  DB 22;  Length 1255;
  Best Local Similarity   97.5%;  Pred. No. 0;
  Matches 1230;  Conservative    4;  Mismatches   20;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||  |    ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoelic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
PI Erickson S, Schwall R, Sliwkowski M;
XX
DR WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoelic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;

Query Match 97.0%; Score 6647; DB 23; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
```

```
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||  |    ||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 5
AAW01111
ID   AAW01111 standard; Protein; 1255 AA.
XX
AC   AAW01111;
XX
DT   01-JAN-1997  (first entry)
XX
DE   HER-2/neu protein.
XX
KW   HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW   breast cancer; ovary cancer; colon cancer; lung cancer;
KW   prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          676..1255
FT                   /label= Intracellular_domain
FT                   /note= "claimed domain, useful for immunisation"
XX
PN   WO9630514-A1.
XX
PD   03-OCT-1996.
XX
PF   28-MAR-1996;   96WO-US01689.
XX
PR   31-MAR-1995;   95US-0414417.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Disis ML;
XX
DR   WPI; 1996-455361/45.
DR   N-PSDB; AAT40739.
XX
PT   DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT   treatment of malignancies with which the HER-2/neu oncogene is
PT   associated
XX
PS   Claim 2; Page 56-61; 71pp; English.
XX
CC   Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC   the product of the HER-2/neo oncogene (see also AAT40739).  The
CC   protein is over-expressed in various cancers, including breast,
CC   ovarian, colon, lung and prostate.  The intracellular domain of the
CC   protein can be used to immunise an animal against a malignancy with
CC   which the oncogene is associated.  The polypeptide can be produced
CC   in transformed host cells for use in immunisation.  Alternatively,
CC   animal cells are transfected in vivo or ex vivo with a viral vector
CC   that directs expression of the polypeptide.
XX
SQ   Sequence   1255 AA;
```

Query Match 96.9%; Score 6641; DB 17; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
```

```
Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                    ||  ||  |   ||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 6
AAW92406
ID   AAW92406 standard; Protein; 1255 AA.
XX
AC   AAW92406;
XX
DT   21-APR-1999  (first entry)
XX
DE   Human HER-2/neu oncogene protein.
XX
KW   HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW   malignancy; treatment; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          676..1255
FT                   /note= "region which elicits immune response"
XX
PN   US5869445-A.
XX
PD   09-FEB-1999.
XX
PF   01-APR-1996;   96US-0625101.
XX
PR   01-APR-1996;   96US-0625101.
PR   17-MAR-1993;   93US-0033644.
PR   12-AUG-1993;   93US-0106112.
PR   31-MAR-1995;   95US-0414417.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Disis ML;
XX
DR   WPI; 1999-152835/13.
DR   N-PSDB; AAX01912.
XX
PT   Use of HER-2/neu polypeptides - for eliciting an immune response to
PT   an HER-2/neu associated malignancy, particularly for treating or
PT   preventing tumours
XX
PS   Claim 3; Column 31-38; 26pp; English.
XX
CC   This sequence represents the human HER-2/neu oncogene protein. A fragment
CC   of this protein is used in a method for eliciting or enhancing an immune
CC   response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC   B cells to produce an immune response to the HER-2/neu protein. The
CC   method can be used for immunisation against a malignancy in which the
CC   HER-2/neu oncogene is associated and in the treatment of an existing
CC   tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match             96.9%;  Score 6641;  DB 20;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db         301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy         361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
```

```
Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                    ||  ||  |   ||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 7
AAB21198
ID   AAB21198 standard; protein; 1255 AA.
XX
AC   AAB21198;
XX
DT   12-JAN-2001  (first entry)
XX
DE   Human HER-2/neu protein.
XX
KW   Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW   breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW   colon cancer.
XX
OS   Homo sapiens.
XX
PN   WO200044899-A1.
XX
PD   03-AUG-2000.
XX
PF   28-JAN-2000; 2000WO-US02164.
XX
PR   29-JAN-1999;   99US-0117976.
XX
PA   (CORI-) CORIXA CORP.
PA   (SMIK ) SMITHKLINE BEECHAM.
XX
PI   Cheever MA,  Gheysen D;
XX
DR   WPI; 2000-505976/45.
DR   N-PSDB; AAB89736.
XX
PT   HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT   useful for vaccinating against breast, ovarian, colon, lung and
PT   prostate cancers -
XX
PS   Claim 52; Fig 7; 128pp; English.
XX
CC   The present sequence is the human HER-2/neu protein. It is a member of
CC   the tyrosine kinase family of receptor-like glycoproteins and shows
CC   homology to the epidermal growth factor receptor (EGFR). It probably
CC   plays a part in cell growth and/or differentiation. The HER-2/neu
CC   gene is an oncogene. An HER-2/neu fusion protein comprising a
CC   HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC   domain may be used to treat or prevent cancer by eliciting or
CC   enhancing an immune response to the HER-2/neu protein. It may be used
CC   to treat malignancies such as breast, ovarian, colon, lung and
CC   prostate cancers, and may be used as an antigen to vaccinate against
CC   these neoplasias.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match           96.9%;  Score 6641;  DB 21;  Length 1255;
  Best Local Similarity 97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                    ||  ||  |  ||||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 8
AAY84780
ID   AAY84780 standard; Protein; 1255 AA.
XX
AC   AAY84780;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW   SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW   tumor cell proliferation; tissue degeneration; arthropathy;
KW   bone resorption; inflammatory disease; degenerative disorder;
KW   wound healing.
XX
OS   Homo sapiens.
XX
PN   WO200020579-A1.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-CA00912.
XX
PR   02-OCT-1998;   98US-0165192.
XX
PA   (UYMC-) UNIV MCMASTER.
XX
PI   Muller WJ,  Siegel PM;
XX
DR   WPI; 2000-303768/26.
DR   N-PSDB; AAA14812.
XX
PT   Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT   erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS   Claim 3; Fig 2; 60pp; English.
XX
CC   The present sequence represents a SPLICE erbB-2 receptor protein. The
CC   protein has an in-frame deletion of 16 amino acids, 2 of which are
CC   conserved cysteine residues, compared to the unspliced protein. The
CC   erbB-2 polynucleotide is used to construct probes for detecting
CC   disorders of cell transformation such as cancer. Antibodies to the
CC   protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC   (e.g. antisense oligonucleotides) which inhibit the expression of
CC   SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC   treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC   for treating conditions involving damaged cells including conditions
CC   in which degeneration of tissue occurs, such as arthropathy, bone
CC   resorption, inflammatory diseases, degenerative disorders of the
CC   central nervous system and wound healing.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match             96.9%;  Score 6641;  DB 21;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
```

```
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720

Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779

Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839

Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899

Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959

Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019

Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079

Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139

Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199

Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259

Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261

Db       1254 PV 1255
```

```
RESULT 9
AAB85458
ID   AAB85458 standard; Protein; 1255 AA.
XX
AC   AAB85458;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human HER-2/neu protein.
XX
KW   Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW   oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS   Homo sapiens.
XX
PN   WO200153463-A2.
XX
PD   26-JUL-2001.
XX
PF   19-JAN-2001; 2001WO-US01850.
XX
PR   21-JAN-2000; 2000US-0177545.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Cheever MA,  Hand-Zimmermann S;
XX
DR   WPI; 2001-476112/51.
DR   N-PSDB; AAH23392.
XX
PT   New antigen-presenting cells, useful as vaccines for eliciting or
PT   enhancing an immune response to HER-2/neu protein, particularly useful
PT   for treating or preventing cancer, e.g. breast cancer  -
XX
PS   Claim 2; Page 41-46; 49pp; English.
XX
CC   The invention provides an isolated antigen-presenting cell, which
CC   expresses at least an immunogenic portion of a polypeptide that produces
CC   an immune response to HER-2/neu protein. The antigen-presenting cells are
CC   useful as vaccines for eliciting or enhancing an immune response to
CC   HER-2/neu protein, particularly in treating or preventing malignancies in
CC   which the HER-2/neu oncogene is associated. Specifically, these are
CC   useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC   colon, lung or prostate cancers. The present sequence represents
CC   the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ   Sequence   1255 AA;

  Query Match           96.9%;  Score 6641;  DB 22;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
```

```
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                ||  ||  |    ||||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 10
AAG88267
ID   AAG88267 standard; Protein; 1255 AA.
XX
AC   AAG88267;
XX
DT   11-SEP-2001  (first entry)
XX
DE   HER2/neu amino acid sequence.
XX
KW   Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW   immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW   tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS   Homo sapiens.
XX
PN   WO200141787-A1.
XX
PD   14-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US33591.
XX
PR   10-DEC-1999;   99US-0458299.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
PI   Keogh E;
XX
DR   WPI; 2001-374995/39.
XX
PT   An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT   cellular immune responses for the prevention and treatment of cancer -
XX
PS   Disclosure; Page 15; 199pp; English.
XX
CC   The present invention describes isolated prepared HER2/neu epitopes (I).
CC   Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC   culture in vitro and binds to a complex of an epitope (I), bound to a
CC   human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC   and a second epitope and the peptide is less than 50 contiguous amino
CC   acids that have 100% identity with a native peptide sequence of HER2/neu;
CC   (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC   excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC   (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC   and immunostimulant activities, and can be used in vaccines. (I), (II)
CC   and (III) are useful for inducing cellular immune responses for the
CC   prevention and treatment of cancer. (I) and (II) are useful for
CC   monitoring or evaluating an immune response to a tumour-associated
CC   antigen when incubated with a T lymphocyte sample form a patient and
CC   detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC   based vaccines mean that immunosuppressive epitopes that may be present
CC   in whole antigens may be avoided. Selected epitopes may be combined to
CC   enhance immunogenicity. The possible pathological side effects caused by
CC   infectious agents or whole protein antigen is eliminated. The vaccine
CC   provides the ability to direct and focus an immune response to multiple
CC   selected antigens from the same pathogen. Epitope-based anti-tumour
CC   vaccines provides the opportunity to combine epitopes derived from
CC   multiple tumour-associated molecules addressing the problem of tumour-
CC   tumour variability and reducing the likelihood of tumour escape due to
CC   antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   1255 AA;
```

Query Match 96.9%; Score 6641; DB 22; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||    ||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX

PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowsert LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6641; DB 23; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
```

```
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720

Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779

Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839

Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899

Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959

Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019

Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079

Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139

Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199

Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259

Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 12
AAE20479
ID   AAE20479 standard; Protein; 1255 AA.
XX
AC   AAE20479;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Human Her-2/neu protein.
XX
KW   Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW   human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1021..1030
FT                   /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN   WO200214503-A2.
XX
PD   21-FEB-2002.
XX
PF   14-AUG-2001; 2001WO-US41733.
XX
PR   14-AUG-2000; 2000US-225152P.
PR   28-SEP-2000; 2000US-236428P.
PR   21-FEB-2001; 2001US-270520P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Hand-zimmermann S,  Cheever MA,  Foy TM,  Lodes MJ,  Kalos MD;
PI   Mcneill PD,  Vedvick TS;
XX
DR   WPI; 2002-280758/32.
DR   N-PSDB; AAD32743.
XX
PT   Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT   prevention and diagnosis of cancer, preferably breast cancer  -
XX
PS   Disclosure; Page 114-117; 129pp; English.
XX
CC   The invention relates to an isolated Her-2/Neu polypeptide composition
CC   effective for eliciting an immune response. The invention is useful for
CC   eliciting an immune response in a patient, where the patient is human
CC   leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC   The composition is useful for the therapy and diagnosis of cancer,
CC   preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC   and other compositions for the diagnosis, prevention and treatment of
CC   human malignancies, for stimulating and/or expanding T cells specific for
CC   Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC   patient. The invention is useful for stimulating a T cell response in a
CC   human patient, as probe or primer for nucleic acid hybridisation, to
CC   selectively form duplex molecules with complementary stretches of the
CC   entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC   length gene from a suitable library, and to direct expression of a
CC   polypeptide in appropriate host cells. The composition is useful in
CC   prophylactic or therapeutic applications and for the treatment of cancer,
CC   preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC   associated malignancies. The invention is useful in gene therapy. The
CC   present sequence is human Her-2/neu protein.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match           96.9%;  Score 6641;  DB 23;  Length 1255;
  Best Local Similarity 97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
```

```
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy        721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                  ||  ||  |    ||||||||||||||||||||||||||||||||||||||||||||
Db        720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy        780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy        840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy        900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy        960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy       1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy       1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy       1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy       1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy       1260 PV 1261
              ||
Db       1254 PV 1255
```

```
RESULT 13
AAM51143
ID   AAM51143 standard; Protein; 1255 AA.
XX
AC   AAM51143;
XX
DT   17-JUN-2002  (first entry)
XX
DE   Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX
KW   Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW   tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..653
FT                   /note= "extracellular domain"
FT   Domain          676..1255
FT                   /note= "intracellular domain"
FT   Domain          990..1255
FT                   /note= "phosphorylation domain"
XX
PN   WO200212341-A2.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-US24283.
XX
PR   03-AUG-2000; 2000US-0632507.
XX
PA   (CORI-) CORIXA CORP.
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Cheever MA,  Gheysen D;
XX
DR   WPI; 2002-241743/29.
DR   N-PSDB; ABA92250.
XX
PT   Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT   or enhancing an immune response to the protein, has Her-2/neu
PT   extracellular domain fused to Her-2/neu intracellular or
PT   phosphorylation domain  -
XX
PS   Claim 68; Fig 7; 141pp; English.
XX
CC   The present sequence is that of human Her-2/neu (p185 glycoprotein
CC   or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC   vaccines.  The Her-2/neu gene is amplified and p185 is overexpressed
CC   in a variety of cancers, including breast, ovarian, colon, lung and
CC   prostate cancer.  Her-2/neu is a member of the tyrosine kinase
CC   family of receptor-like glycoproteins.  It comprises an extracellular
CC   domain with homology to the epidermal growth factor receptor
CC   (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC   intracellular domain that also shows homology to EGFR.  Its
CC   overexpression correlates with a poor prognosis in breast and
CC   ovarian cancers.  The invention provides Her-2/neu fusion
CC   proteins, nucleic acids encoding them, viral vectors, and vaccines
CC   comprising the fusion proteins or nucleic acid molecules.  In
CC   preferred fusion proteins, the extracellular domain of a Her-2/neu
CC   protein is fused to a Her-2/neu intracellular domain or
CC   phosphorylation domain (or its DeltaPD fragment).  An immune
CC   response to Her-2/neu protein is elicited or enhanced by
CC   administering the fusion protein in the form of a vaccine, or by
CC   transfecting cells of an animal ex vivo with a nucleic acid
CC   encoding the fusion protein, and delivering the transfected cells
CC   to the animal.  The fusion proteins, nucleic acids, and isolated
CC   specific T-cells are useful for inhibiting the development of a
CC   cancer, especially breast, ovarian, colon, lung or prostate cancer
CC   in a patient.  T cells that specifically react with a Her-2/neu
CC   fusion protein can be used to remove tumour cells from a sample in
CC   order to inhibit the development of cancer in a patient.
XX
SQ   Sequence   1255 AA;
```

```
  Query Match           96.9%;  Score 6641;  DB 23;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
```

```
Qy        61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy       121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy       181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy       361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy       661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy       721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
                 ||  ||    |||||||||||||||||||||||||||||||||||||||||||||
Db       720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy       780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy       840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy       900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy       960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy      1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy      1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy      1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy      1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy      1260 PV 1261
             ||
Db      1254 PV 1255
```

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide -
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

```
  Query Match             96.9%;  Score 6641;  DB 23;  Length 1255;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1228;  Conservative    5;  Mismatches   21;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
```

```
Db        61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy       121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db       121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy       181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db       181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360

Db       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy       361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db       361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Db       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Db       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660

Db       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy       661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720

Db       661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy       721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779

Db       720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy       780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839

Db       774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy       840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899

Db       834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy       900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959

Db       894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy       960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019

Db       954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy      1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079

Db      1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy      1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139

Db      1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy      1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199

Db      1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy      1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259

Db      1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy      1260 PV 1261

Db      1254 PV 1255
```

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells. such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isolelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

```
  Query Match             96.3%;  Score 6598;  DB 14;  Length 1433;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1221;  Conservative    7;  Mismatches   26;  Indels    8;  Gaps    2;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQVCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEHCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
```

```
Qy       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db       241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
Db       301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy       361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
Db       361 IQEFAGCKRIFGSLAFLPESFDGDPASNTAPLQPEHLQVFETLEQITGYLYISAWPDSLP 420

Qy       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
Db       421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLSFVHTV 480

Qy       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db       481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy       541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
Db       541 VEECRVLQGLPREYVNASHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
Db       601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy       661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKEFNN 720
Db       661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETE- 719

Qy       721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 779
Db       720 ------LRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 773

Qy       780 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 839
Db       774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833

Qy       840 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 899
Db       834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893

Qy       900 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 959
Db       894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953

Qy       960 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1019
Db       954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013

Qy      1020 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1079
Db      1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073

Qy      1080 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133

Qy      1140 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1199
Db      1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1193

Qy      1200 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1259
Db      1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253

Qy      1260 PV 1261
Db      1254 PV 1255
```

Search completed: July 22, 2003, 09:16:49
Job time : 42.1104 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-710-730-14
Perfect score: 6813
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6602	96.9	1255	1	A24571	protein-tyrosine k
2	5805	85.2	1260	1	TVRTNU	protein-tyrosine k
3	5795.5	85.1	1254	2	I48161	p-185 precursor -
4	3035	44.5	1210	1	GQHUE	epidermal growth f
5	3006	44.1	1210	2	A53183	epidermal growth f
6	2983.5	43.8	1223	1	TVCHLV	epidermal growth f
7	2852.5	41.9	1308	2	A47253	epidermal growth f
8	2559	37.6	1166	1	S06142	protein-tyrosine k
9	2307.5	33.9	1342	2	A36223	kinase-related tra
10	2221.5	32.6	1339	2	JC4387	epidermal growth f
11	1655.5	24.3	698	1	TVFVLV	protein-tyrosine k
12	1592	23.4	604	1	TVYUH	protein-tyrosine k
13	1541	22.6	544	2	S35745	protein-tyrosine k
14	1534	22.5	545	2	S00727	kinase-related tra
15	1517	22.3	540	2	B44776	protein-tyrosine k
16	1515	22.2	540	1	TVFVEB	protein-tyrosine k
17	1514.5	22.2	1330	1	GQFFE	epidermal growth f
18	1509	22.1	644	2	A36325	epidermal growth f
19	1217	17.9	1323	2	E88257	protein let-23 [im
20	1217	17.9	1374	2	S70712	protein-tyrosine k
21	1131	16.6	1369	2	S70713	protein-tyrosine k
22	1126	16.5	527	2	A42032	epidermal growth f
23	1114	16.4	1717	1	A45558	epidermal growth f
24	975.5	14.3	843	2	A27131	epidermal growth f
25	806.5	11.8	346	2	S13807	protein-tyrosine k
26	754.5	11.1	311	2	S13808	protein-tyrosine k
27	693	10.2	1363	2	T43220	insulin-like growt
28	657	9.6	1382	1	INHUR	insulin receptor p
29	656.5	9.6	1372	2	A34157	insulin receptor p
30	653	9.6	1383	2	A36080	insulin receptor p
31	643.5	9.4	1300	2	A36502	insulin receptor-r
32	632	9.3	1477	2	T18534	protein-tyrosine k
33	630	9.2	1607	2	T43212	insulin-like growt
34	629	9.2	1268	2	B36502	insulin receptor-r
35	602	8.8	1367	1	IGHUR1	insulin-like growt
36	583	8.6	1371	2	A33837	insulin-like growt
37	581.5	8.5	2148	1	A56081	insulin receptor -
38	580	8.5	1114	1	S05582	protein-tyrosine k
39	579	8.5	987	2	A54092	protein-tyrosine k
40	575	8.4	2101	2	S57245	insulin receptor (
41	569	8.4	1390	2	T30346	insulin receptor -
42	559	8.2	987	2	I48652	mouse developmenta
43	551.5	8.1	952	2	I50612	protein-tyrosine k
44	548.5	8.1	801	4	TVHURE	transforming prote
45	548.5	8.1	984	2	A39753	protein-tyrosine k

ALIGNMENTS

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito,
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal grow
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epi
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chr
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517,'RALL',522,'S',524-654,'V',656-1169,'A',1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:g183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

```
A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TMM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

  Query Match             96.9%;  Score 6602;  DB 1;  Length 1255;
  Best Local Similarity   96.9%;  Pred. No. 4.6e-277;
  Matches 1221;  Conservative   11;  Mismatches   18;  Indels   10;  Gaps    3;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||

Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANKEILDEAYVMA 775
              |||||||||||||    ||   : :|      |:  : ::|||||||||||||||||||||
Db        721 RKVKVLGSGAFG----TVYKGIWIPDGENVKIPVA-IKVLRENTSPKANKEILDEAYVMA 775

Qy        776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835

Qy        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895

Qy        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955

Qy        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1015

Qy       1016 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1075

Qy       1076 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135

Qy       1136 QPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPE 1195
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1136 QPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPE 1195

Qy       1196 YLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 YLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255


RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protei
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Coh
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
```

```
:8-680/Domain: transmembrane #status predicted <TMN>
:3-988/Domain: protein kinase homology <KIN>
:1-739/Region: protein kinase ATP-binding motif
:,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
:1/Binding site: phosphate (Thr) (covalent) #status predicted
:8/Active site: Lys #status predicted
32,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

uery Match             85.2%;  Score 5805;  DB 1;  Length 1260;
est Local Similarity   85.3%;  Pred. No. 9.4e-243;
atches 1076;  Conservative   58;  Mismatches  116;  Indels   12;  Gaps    5;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||
Db          4 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 63

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| ||||||||||||||||:|||||||::||||||||||||||||| ||||||||
Db         64 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 123

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||||: |||||||| :||||:| |||||
Db        124 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 183

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |  |||||||| ||:| || : ||||| |||| || |:|  ||||||| |||||||||
Db        184 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 243

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        244 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 303

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||||||||||||||||::::
Db        304 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 363

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
              || |||||||||||||||||||:|  |||:|||||||||||||||||||||||||||||
Db        364 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 423

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              ||||||||::||||||:|||||||||||||  |||||||||||||||| | |||||||
Db        424 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 483

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||: ||||:: ||  || |: ||| |||||||||||||| |||||
Db        484 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 543

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : ||||||||||| | |||| ||||| |||||||    |||
Db        544 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 603

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAV 658
              |||||||||||||||:||||| |||||||||||||||::|||||||||:| |:: |
Db        604 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRASPVTFIIATV 663

Qy        659 VGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 718
              |:|| ::| || ||||||||:|||||||||||||||||||||||||||||||||||||||
Db        664 EGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 723

Qy        719 ELRKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANKEILDEAYV 773
              |||||||||||||    ||   : :|     |:  :  ::|||||||||||||||||||
Db        724 ELRKVKVLGSGAFG----TVYKGIWIPDGENVKIPVA-IKVLRENTSPKANKEILDEAYV 778

Qy        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833
              ||||||||||||||||||||||||||||||||||||:|||||||||||:|||||||||||
Db        779 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGM 838

Qy        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 898

Qy        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 958

Qy        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013
              |||||||||||||||||||||||||||||||||||||||||||||:||:|||||||||||
Db        959 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDDD 1018

Qy       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073
              |||||||||||||||||| ||| || |  | ||||||||||:|||||||||  |||
Db       1019 MGDLVDAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPPRS 1078

Qy       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133
              |||||||||||||||| ||  |||||| || ||||||||||:||| |||||||| |
Db       1079 PLAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPLAC 1138

Qy       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1193
              ||||||||| :|:|||||  |||||  |||||||| ||||||||||||||||||||||||
Db       1139 SPQPEYVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1198

Qy       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253
              |||| |: | |  ||| |||||||||||||||:  |:| ||| |:|||||||||||||
Db       1199 PEYLVPREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGLDV 1258

Qy       1254 PV 1255
              ||
Db       1259 PV 1260
```

RESULT 3
I48161
p-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MUID:94193007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

```
  Query Match             85.1%;  Score 5795.5;  DB 2;  Length 1254;
  Best Local Similarity   84.9%;  Pred. No. 2.4e-242;
  Matches 1070;  Conservative   67;  Mismatches  112;  Indels   11;  Gaps    4;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              |||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:||||:||| ||||||||||||||||||||||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTVLWKDVFRKNNQLA 180
              |||:|  |  || :| |||||||||||||||||||||||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||||  ||:| || : ||| | ||||:|| |:         :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300
```

```
 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
 301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
 361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
 421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

 541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
 541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
 601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

 661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

 721 RKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANKEILDEAYVMA 775
 721 RKVKVLGSGAFG----TVYKGIWIPDGENVKIPVA-IKVLRENTSPKANKEILDEAYVMA 775

 776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835
 776 GLGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSY 835

 836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895
 836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALESIL 895

 896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955
 896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955

 956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1015
 956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDMG 1015

1016 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1075
1016 DLVDAEEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRSPL 1075

1076 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
1076 APSEGAGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLACSP 1135

1136 QPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPE 1195
1136 QPEYVNQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVENPE 1195

1196 YLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
1196 YLVPRGGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDVPV 1254
```

```
LT 4
E
ermal growth factor receptor precursor - human
ntains: protein-tyrosine kinase (EC 2.7.1.112) erbB
ecies: Homo sapiens (man)
te: 15-Nov-1984 #sequence_revision 27-Nov-1985 #text_change 11-Jun-1999
cession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33
```

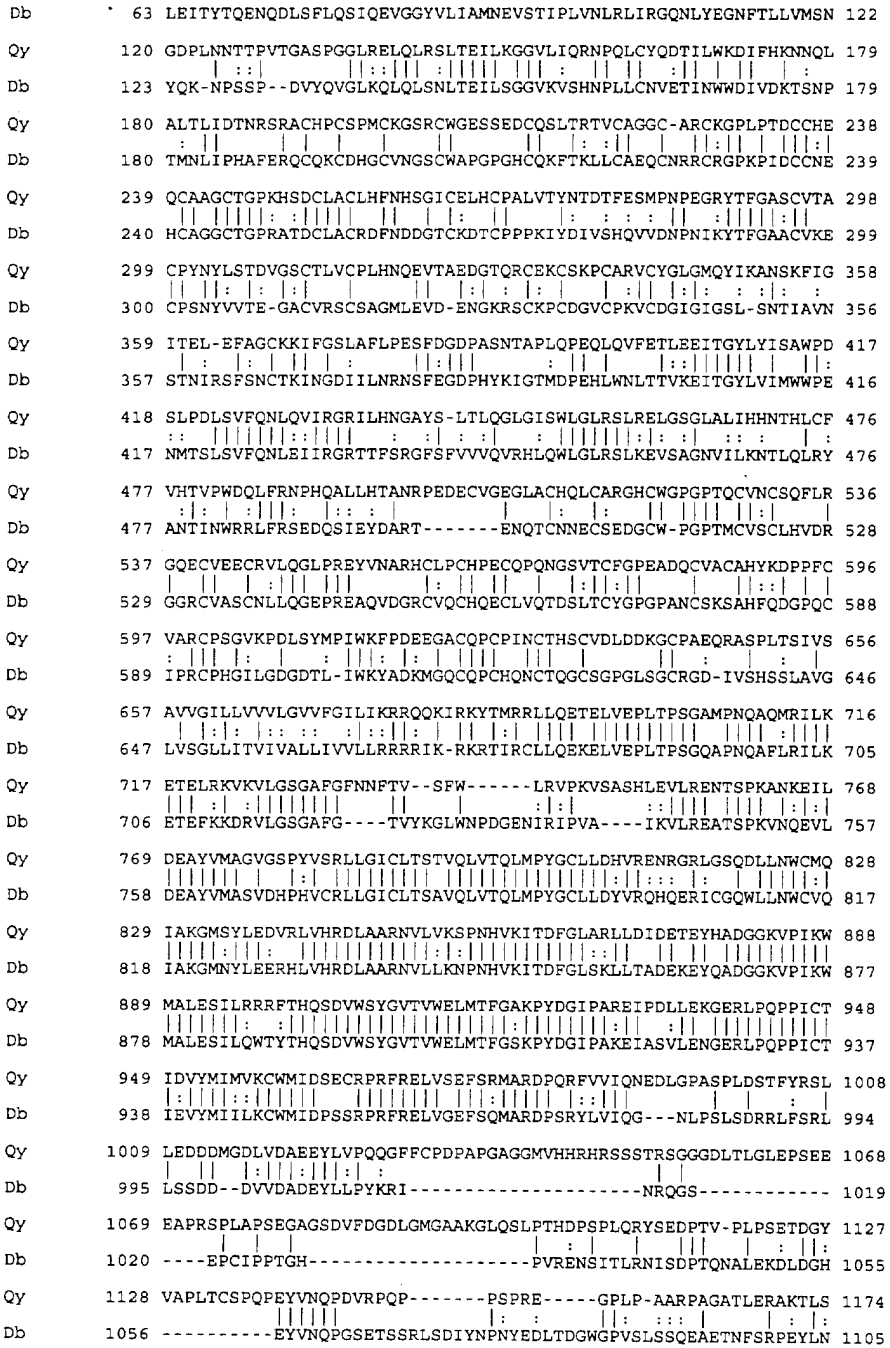
R;Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Ya
rg, P.H.
Nature 309, 418-425, 1984
A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210 <ULL>
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epidermal grov
A;Reference number: A25772; MUID:85270438; PMID:2991899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identification of
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript terminati
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Merl
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs c
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 'RCAWRRA',150-187,'KSVIQAV',195,'M',197,'A',199-222,'S',224-304,'RA',307-321,
',798-799,'TD',802-811,'R',813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF recept
R;Lin, C.R.; Chen, W.S.; Kruiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.M
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplification a
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30,'S',32-51;454-467 <WEB>
R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A;Title: Identification of residues in the nucleotide binding site of the epidermal growt
A;Reference number: A60143; MUID:85182650; PMID:2985580
A;Accession: A60143
A;Molecule type: protein
A;Residues: 740-744,'X',746-747 <RUS>
R;Mroczkowski, B.; Mosig, G.; Cohen, S.
```
```

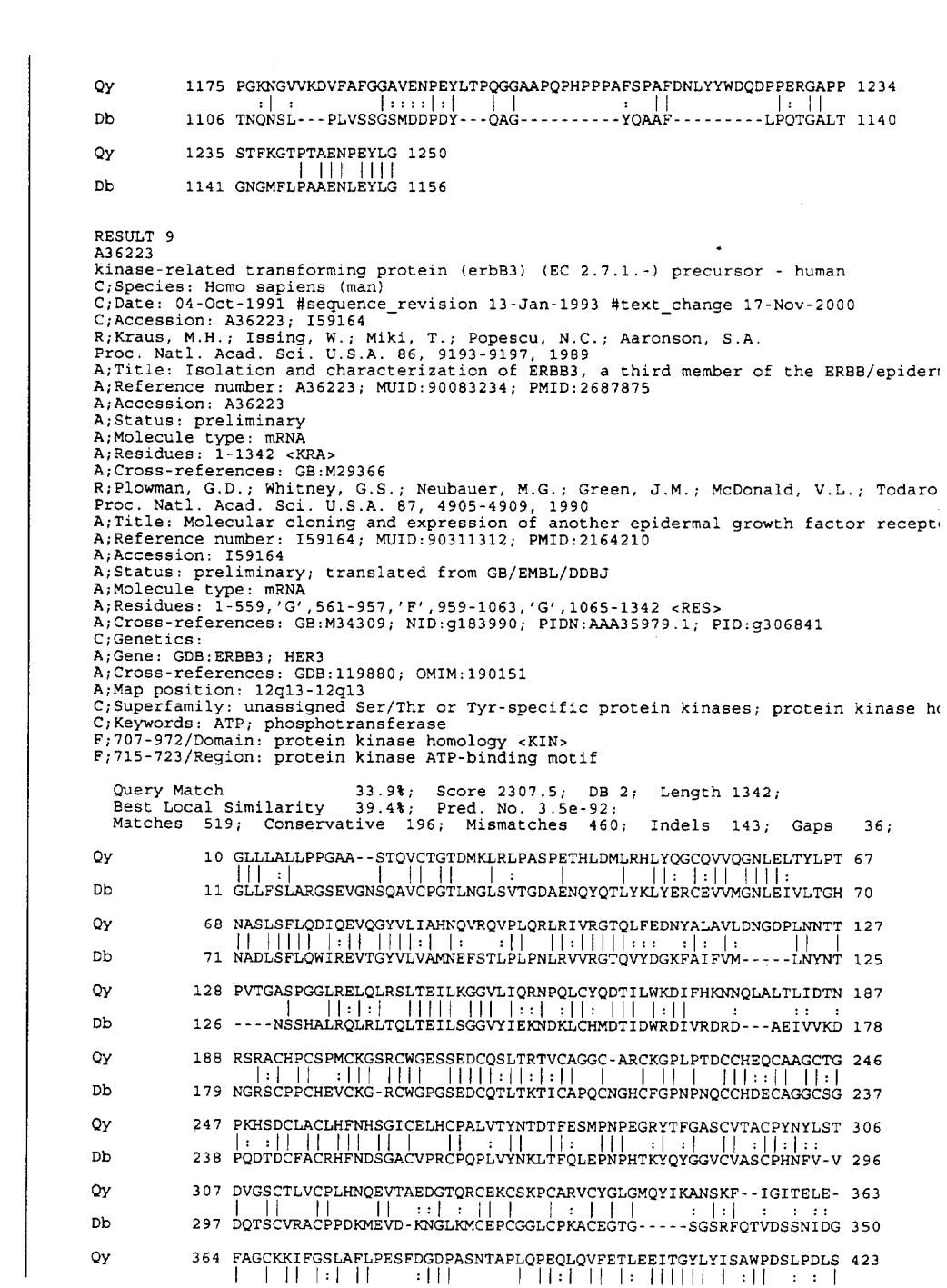
```
Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and supercd
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity
A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain r
A;Reference number: A33331; MUID:90003233; PMID:2790960
A;Contents: annotation; internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:120610; OMIM:131550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphc
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1210/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <EE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <EE2>
F;646-668/Domain: transmembrane #status predicted <TMM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-726/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F;745/Active site: Lys #status experimental

  Query Match           44.5%;  Score 3035;  DB 1;  Length 1210;
  Best Local Similarity  48.7%;  Pred. No. 1.5e-123;
  Matches  619;  Conservative  176;  Mismatches  359;  Indels  116;  Gaps   25;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
Db        127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy        189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
Db        184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy        248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
Db        244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy        308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
Db        304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy        367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
Db        362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy        427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
Db        422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy        487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
Db        482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCKL 541

Qy        547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606

Db        542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy        607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG---ILL 663
Db        602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLLL 658

Qy        664 VVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKV 723
Db        659 VVALGIG---LFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKI 715

Qy        724 KVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSPKANKEILDEAYVMAGVGS 779
Db        716 KVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDN 771

Qy        780 PYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDV 839
Db        772 PHVCRLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDR 831

Qy        840 RLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRF 899
Db        832 RLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIY 891

Qy        900 THQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCW 959
Db        892 THQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCW 951

Qy        960 MIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLV 1018
Db        952 MIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVV 1011

Qy       1019 DAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPS 1078
Db       1012 DADEYLIPQQGFF----------------------------------SSPSTSRTPLLSS 1037

Qy       1079 EGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQ 1136
Db       1038 LSATSN--NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------PV 1089

Qy       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1196
Db       1090 PEYINQ-SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNPEY 1138

Qy       1197 L-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKG 1239
Db       1139 LNTVQ---------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKG 1189

Qy       1240 TPTAENPEYL 1249
Db       1190 S-TAENAEYL 1198


RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C;Accession: A53183; A43818; S24942; A28941; S45325; I49643
R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF recepto
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:U03425
R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding
A;Reference number: A43818; MUID:91232866; PMID:2030916
A;Accession: A43818
A;Molecule type: mRNA
A;Residues: 1-714 <AVI>
A;Cross-references: GB:X59698
```

```
R;Eisinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A;Reference number: S24942
A;Accession: S24942
A;Molecule type: mRNA
A;Residues: 969-971,'K',973-1115,'D' <EIS>
A;Cross-references: EMBL:Z12608
R;Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A;Reference number: A28941; MUID:88330814; PMID:3138233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971,'K',973-1210 <VER>
A;Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A;Reference number: I49643; MUID:93126380; PMID:7678348
A;Accession: I49643
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 12-20,22-132 <RES>
A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
C;Genetics:
A;Gene: EGFR
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphopro
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TMM>
F;712-977/Domain: protein kinase homology <KIN>
F;720-728/Region: protein kinase ATP-binding motif
F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

  Query Match             44.1%;  Score 3006;  DB 2;  Length 1210;
  Best Local Similarity   48.6%;  Pred. No. 2.7e-122;
  Matches  621;  Conservative  169;  Mismatches  367;  Indels  120;  Gaps   27;

Qy          11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
               ||  ||    | |     :|| ||  :|    : | |    |: :|  |:|| ||||:||:  |
Db          14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy          69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
                 ||||: |||| |||||| | | ::||: |:|:||  |:|: ||||:| |
Db          74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy         129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
                 |  :  |||||  :|:| ||| | |    || ||  ||| |:||    |  |  : |
Db         125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy         185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
                   :   :|  | | |     |||   |:|| ||: :||  |: ||:|  |:|||| |||||
Db         181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy         244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
               ||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:||  || ||
Db         240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy         304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
               : || |||   |    ||  ||| ::|:||  || :|| |:|:   | ::  |  | ::
Db         300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy         364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
                |  |  | | |  || :| ||  : | || | :|:: :|::||||:| | ||||:  ||
Db         358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy         423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
                 |:||::|||   :| :|| : || |: ||||||:|:  |  :|  | :||: :|: |
Db         418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy         483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
                :||  |:|      || | :|      |: ||:   |||| |  ||:|    ||:||||
Db         478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy         543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
               :|:|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||  ||:
Db         538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy         603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
               |:  : : : :||: |    |  | |||: |    :||         : || : :|| |
Db         598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy         663 LVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELR 721
               | :|: |  || |  ||:  :|| |:|||||| |||||||||  |||| :|||||||  :
Db         657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFK 715

Qy         722 KVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSPKANKEILDEAYVMAGV 777
               |:|||||||||     ||    | :|:      :  |    ||| ||||||||||||||||| |
Db         716 KIKVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASV 771

Qy         778 GSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLE 837
                 :|:| |||||||||||||:|||||||||||:|||::   :||| |||||:||||||:|||
Db         772 DNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLE 831

Qy         838 DVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRR 897
               | |||||||||||||||:| ||||||||||:||  :| ||||:|||||||||||||| |
Db         832 DRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHR 891

Qy         898 RFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVK 957
                :|||||||||||||||||||:|||||||| :|  :|||||||||||||||||||||||
Db         892 IYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVK 951

Qy         958 CWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGD 1016
               |||||:: ||:|||: |||:|||||||:|||  :|  :    || || |||:|::::|| |
Db         952 CWMIDADSRPKFRELILEFSQMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMED 1011

Qy        1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076
               :|||:|||:||||||                    | ||                 |:||
Db        1012 VVDADEYLIPQQGFF----------------NSPST-----------------SRTPLL 1037

Qy        1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCS 1134
                 |  | |:    :  :           |    : : ||||| ||| :  :   | ::
Db        1038 SSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------ 1087

Qy        1135 PQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENP 1194
               | ||||||  |  :|    : |:   :|           :||::   ::      || ||
Db        1088 PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNP 1136

Qy        1195 EYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTF 1237
               |||  |  |          |   |  |::   | |          :|       |:   |   |
Db        1137 EYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIF 1187

Qy        1238 KGTPTAENPEYLGLDVP 1254
               || ||||| ||| :  |
Db        1188 KG-PTAENAEYLRVAPP 1203


RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Gallus gallus (chicken)
C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C;Accession: A27720; A00643
```

```
R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A;Reference number: A27720; MUID:88261272; PMID:3260329
A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
pecific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TMM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

  Query Match             43.8%;  Score 2983.5;  DB 1;  Length 1223;
  Best Local Similarity   47.5%;  Pred. No. 2.6e-121;
  Matches  619;  Conservative  174;  Mismatches  354;  Indels  155;  Gaps   29;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIV 655
Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 SAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRI 714
Db        657 AGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRI 715

Qy        715 LKETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSPKANKEILDE 770
Db        716 LKETEFKKVKVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDE 771

Qy        771 AYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIA 830
Db        772 AYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNWCVQIA 831

Qy        831 KGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMA 890
Db        832 KGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPIKWMA 891

Qy        891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950
Db        892 LESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPPICTID 951

Qy        951 VYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLL 1009
Db        952 VYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPTDSKFYRTLM 1011

Qy       1010 EDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEE 1069
Db       1012 EEEDMEDIVDADEYLVPHQGFF----------------NSPST----------------- 1038

Qy       1070 APRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET 1124
Db       1039 -SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSDPTGNFLEES 1089

Qy       1125 --DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVK 1182
Db       1090 IDDGFL------PAPEYVNQ--LMPKKPS----------------------TAMVQNQIYN 1120

Qy       1183 DVF---------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQ-- 1225
Db       1121 NISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFESSPYWIQSG 1172

Qy       1226 ------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db       1173 NHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 1214

RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
```

```
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

  Query Match             41.9%;  Score 2852.5;  DB 2;  Length 1308;
  Best Local Similarity   44.1%;  Pred. No. 1.2e-115;
  Matches  596;  Conservative  187;  Mismatches  385;  Indels  183;  Gaps   31;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |:  |      |  || ||: ||   :  |     ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              :  |  |||| : ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::|||| ||| : :|  ||| ||| |:||        |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||   ||  ||:   |  :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||    ||  |:| : |: |:  | : | |:|   : :         :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :    : ||:| || |: |||:| | | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||   :| ::   ||  : |  ||: |: ||:   |||||| ||::| :| ||: |:|
Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              | :  |  ||: |   |: | |:|:  ::| :|| ||  | |  |:|:|| | || :||
Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PAEQRASPL 651
              |::   |:  |:|: | :  | ||  ||| |       |            |   | :||
Db        596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQHAR-TPL 652

Qy        652 TSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710
                | : |: |: ::|::|: | : ::|:  | :|  :|| | ||||||||||||  ||||
Db        653 --IAAGVIGGLFILVIVGLTFAVYVRRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNQA 708

Qy        711 QMRILKETELRKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANK 765
              |:|||||||::||||||||||     ||   : ||     |:  : :::| | | ||||
Db        709 QLRILKETELKRVKVLGSGAFG----TVYKGIWVPEGETVKIPVA-IKILNETTGPKANV 763

Qy        766 EILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNW 825
              | :||| :|| :  |:: ||||:|| |:|||||||:||||::| |::  :||| ||||
Db        764 EFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLLNW 823

Qy        826 CMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVP 885
              |:|||||| |||: |||||||||||||||||||||||||||||||: || ||:|||||:|
Db        824 CVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKMP 883

Qy        886 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPP 945
              |||||||  |  |:|||||||||||||:|||||||  ||||||| |||||||||||||||
Db        884 IKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQPP 943

Qy        946 ICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDSTF 1004
              |||||||:|||||||:: ||:|:|| :|||||||||||::||| :|  :   || || |
Db        944 ICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDSKF 1003

Qy       1005 YRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLE 1064
              :::||:::|: |::|||||||| | |  | |       :   | |  | ||      :|
Db       1004 FQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIGHS 1051

Qy       1065 PSEEEAPRS--------------------------PLAP-SEGAGSDVFDGDLGMGAA 1095
              |     | |                          | || ::|| :::||      |
Db       1052 PPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNGTL 1111

Qy       1096 KGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQ 1148
              :   :    : |  |||| ||||  |        : :||: |:   |: ||:|  :
Db       1112 RKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE---- 1167

Qy       1149 PPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPH 1208
                | |  : |               ||| ::       |::|||      |
Db       1168 -----ENPFVSRR--------------KNGDLQ-------ALDNPEYHNASNG------ 1194

Qy       1209 PPPA--------------------------------FSPAFDNLYYWDQDPPERGA-- 1232
              || |                                ||||  ||:   | |
Db       1195 PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRSTLQ 1254

Qy       1233 PPSTFKGTPT--------------AENPEYL 1249
                |   :   |               |||||||
Db       1255 HPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transfo
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueler, W.; Raulf, F.; Telling, A.; Ro
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueler, W.; Schartl, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025,'N',1027-1098,'A',1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

  Query Match             37.6%;  Score 2559;  DB 1;  Length 1166;
  Best Local Similarity   43.9%;  Pred. No. 4.6e-103;
  Matches  560;  Conservative  168;  Mismatches  392;  Indels  156;  Gaps   34;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              |||  |||  |      ||    :|| ||  :: :    : |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||   |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: | |: |
```





RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C;Accession: A36223; I59164
R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epider
A;Reference number: A36223; MUID:90083234; PMID:2687875
A;Accession: A36223
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1342 <KRA>
A;Cross-references: GB:M29366
R;Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A;Title: Molecular cloning and expression of another epidermal growth factor recept
A;Reference number: I59164; MUID:90311312; PMID:2164210
A;Accession: I59164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-559,'G',561-957,'F',959-1063,'G',1065-1342 <RES>
A;Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C;Genetics:
A;Gene: GDB:ERBB3; HER3
A;Cross-references: GDB:119880; OMIM:190151
A;Map position: 12q13-12q13
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; phosphotransferase
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

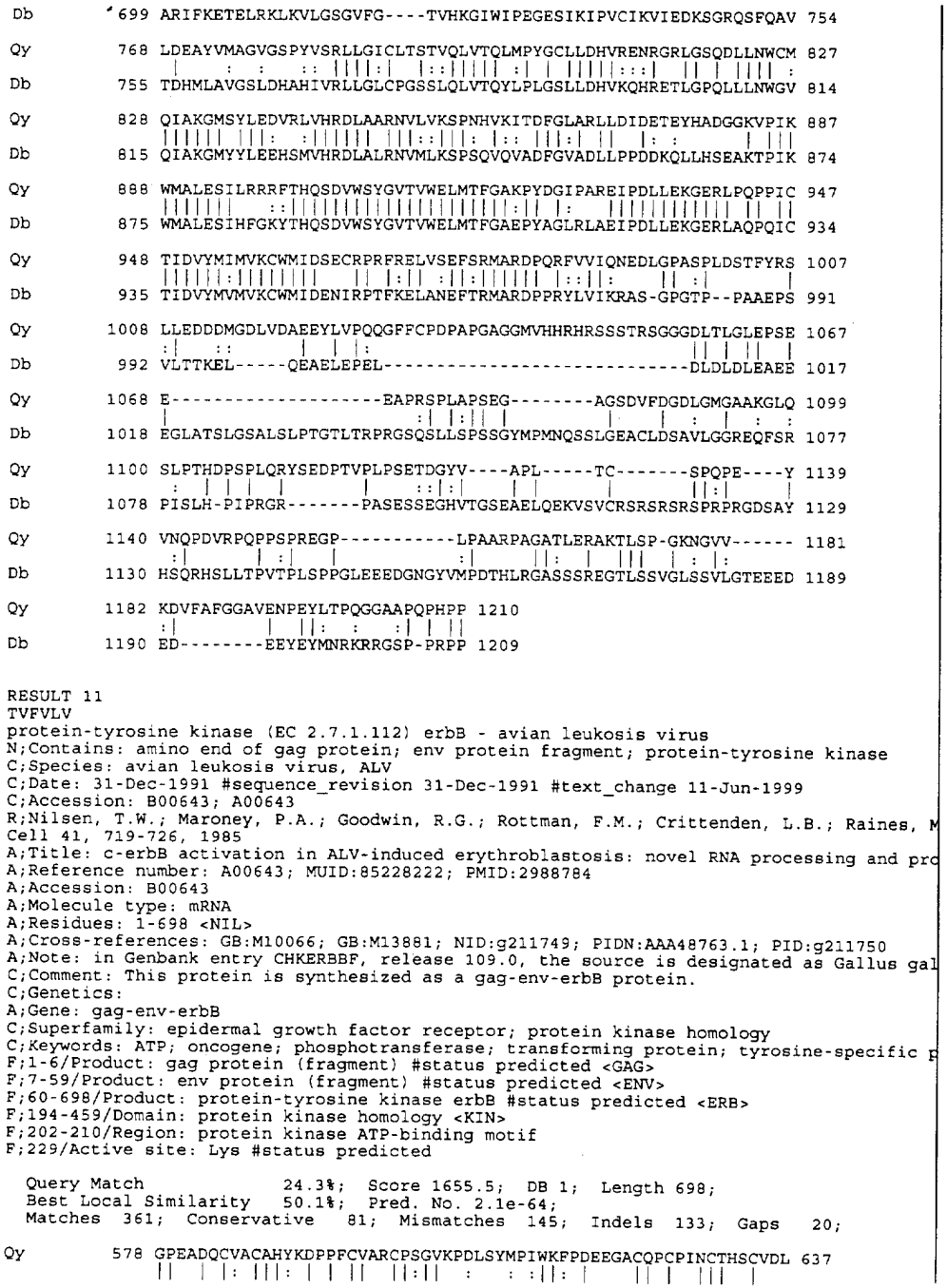
Query Match 33.9%; Score 2307.5; DB 2; Length 1342;
Best Local Similarity 39.4%; Pred. No. 3.5e-92;
Matches 519; Conservative 196; Mismatches 460; Indels 143; Gaps 36;

```
Db        351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVFRTVREITGYLNIQSWPPHMHNFS 410

Qy        424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482

Db        411 VFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNW 470

Qy        483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECV 541

Db        471 TKVLRGPTEERLDIKHNRPRRDCVAEGKVCDPLCSSGGCWGPGPGQCLSCRNYSRGGVCV 530

Qy        542 EECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCP 601

Db        531 THCNFLNGEPREFAHEAECFSCHPECQPMEGTATCNGSGSDTCAQCAHFRDGPHCVSSCP 590

Qy        602 SGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRA----SPLTSIVSA 657

Db        591 HGVLG--AKGPIYKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLIGKTHLTMALTV 648

Qy        658 VVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716

Db        649 IAG--LVVIFMMLGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705

Qy        717 ETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASH----LEVLRENTSPKANKEILDEAY 772

Db        706 ETELRKLKVLGSGVFG----TVHKGVWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHML 761

Qy        773 VMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKG 832

Db        762 AIGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVRQHRGALGPQLLLNWGVQIAKG 821

Qy        833 MSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALE 892

Db        822 MYYLEEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKWMALE 881

Qy        893 SILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVY 952

Db        882 SIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEVPDLLEKGERLAQPQICTIDVY 941

Qy        953 MIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPA---SPLDSTFYRSLL 1009

Db        942 MVMVKCWMIDENIRPTEKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKL 1000

Qy       1010 EDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP-SEE 1068

Db       1001 EEVELEPELDLDLDLEAEED------------------NLATTTLGSALSLPVGTLNRP 1041

Qy       1069 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHD-PSPLQRYSEDPTVPLP------ 1121

Db       1042 RGSQSLLSPSSGY-MPMNQGNLGESCQESAVSGSSERCPRPVSLH------PMPRGCLAS 1094

Qy       1122 SETDGYVA----------------PLTCSPQPE----YVNQPDVRPQPPSPREGP----- 1156

Db       1095 ESSEGHVTGSEAELQEKVSMCRSRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPPGLEEE 1154

Qy       1157 ------LPAARPAGATLERAKTLSP-GKNGVV------KDVFAFGGAVENPEYLTPQGGA 1203

Db       1155 DVNGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEEEDED--------EEYEYMNRRRRH 1206

Qy       1204 APQPHPPPAFSPAFDNLYYWD-----------QDPPERGAPPSTFKGTPTAENPEYL 1249

Db       1207 SP-PHPPRPSSLEELGYEYMDVGSDLSASLGSTQSCPLHPVPIMPTAGTTPDEDYEYM 1263
```

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res
C;Comment: This protein is a functional heregulin receptor that transduces signals
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TMM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Ty

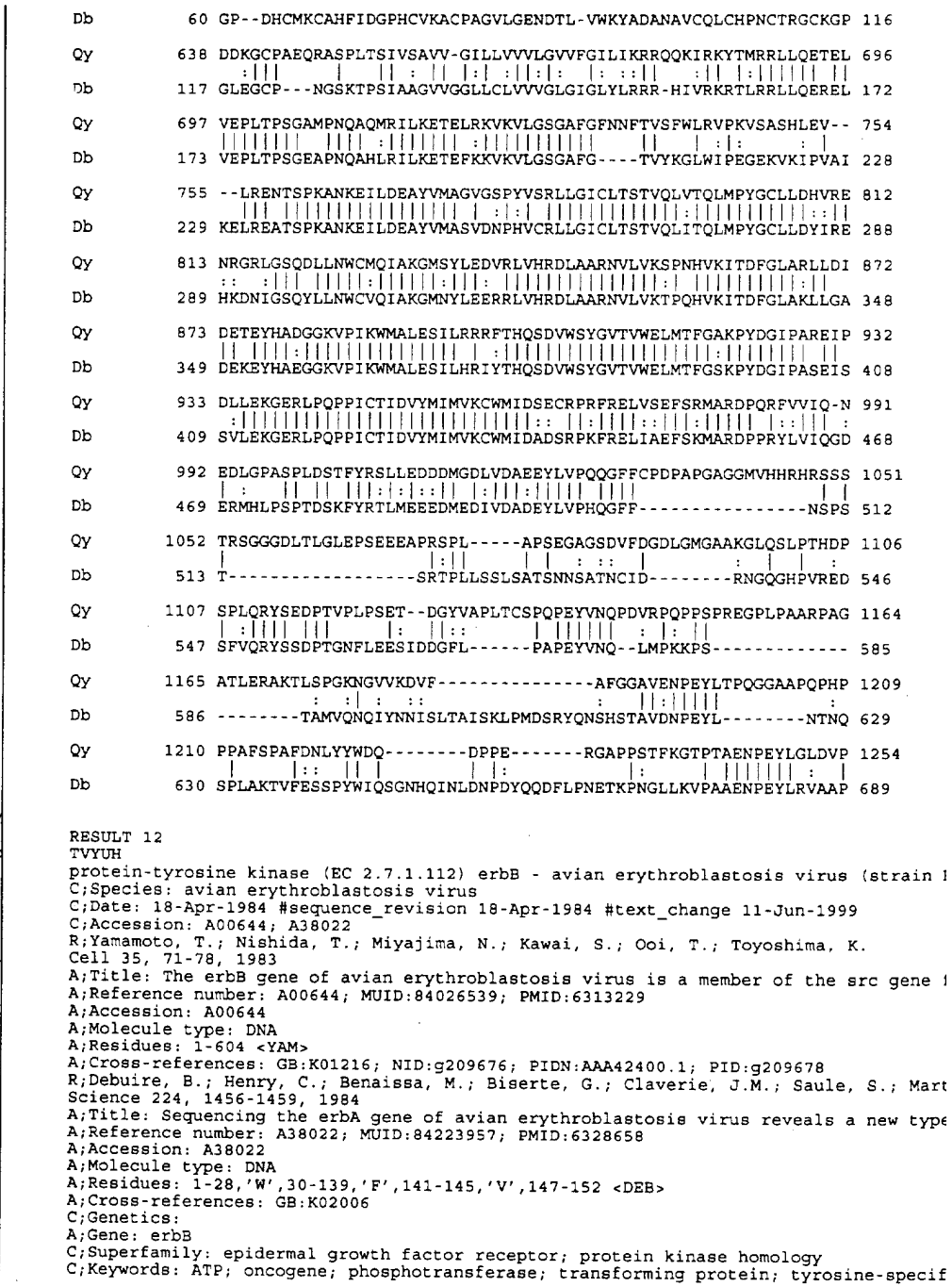
  Query Match 32.6%; Score 2221.5; DB 2; Length 1339;
  Best Local Similarity 39.5%; Pred. No. 1.8e-88;
  Matches 509; Conservative 173; Mismatches 440; Indels 167; Gaps 37;

```
Qy          3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59

Db          7 LQVLC----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLYKLYEKCEVVMGN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119

Db         63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNEFSVLPLPNLRVVRGTQVYDGKFAIFVM-- 120

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179

Db        121 ---LNYNT----NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRVR--- 170

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238

Db        171 GAEIVVKNNGANCPPCHEVCKG-RCWGPGPDDCQILTKTICAPQCNGRCFGPNPNQCCHD 229

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298

Db        230 ECAGGCSGPQDTDCFACRRFNDSGACVPRCPEPLVYNKLTFQLEPNPHTKYQYGGVCVAS 289

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGL--GMQYIKANSKF 356

Db        290 CPHNFV-VDQTFCVRACPPDKMEVD-KHGLKMCEPCGGLCPKACEGTGSGSRYQTVDSSN 347

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416

Db        348 ID----GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLNVFRTVREITGYLNIQSWP 403

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLC 475

Db        404 PHMHNFSVFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLC 463

Qy        476 FVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQF 534

Db        464 YHHSLNWTRLLRGPSEERLDIKYDRPLGECLAEGKVCDPLCSSGGCWGPAPGQCLSCRNY 523

Qy        535 LRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPP 594

Db        524 SREGVCVTHCNFLQGEPREFVHEAQCFSCHPECLPMEGTSTYNGSGSDACARCAHFRDGP 583

Qy        595 FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSC--VDLDDKGCPAEQRASPLT 652

Db        584 HCVNSCPHGILG--AKGPIYKYPDAQNECRPCHENCTQGCNGPELQDCLGQAEVLMSKPH 641

Qy        653 SIVSAVVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGAMPNQAQ 711

Db        642 LVIAVTVG--LAVILMILGGSFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVL 698

Qy        712 MRILKETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASH----LEVLRENTSPKANKEI 767
```



RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

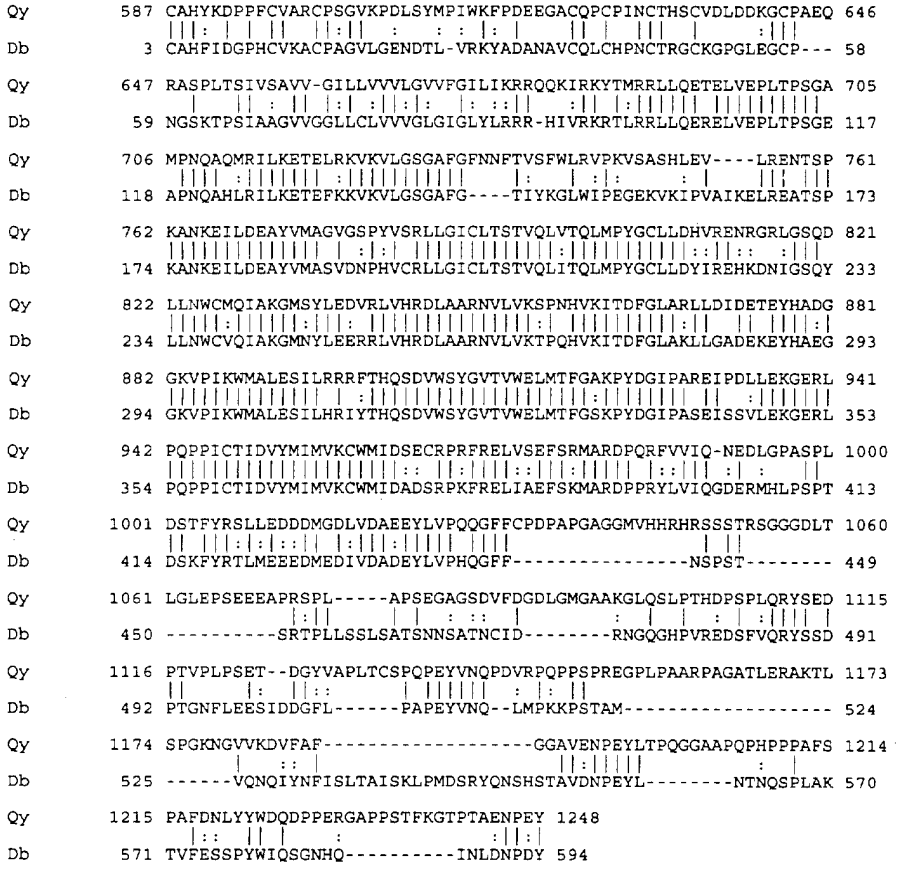
Query Match 24.3%; Score 1655.5; DB 1; Length 698;
Best Local Similarity 50.1%; Pred. No. 2.1e-64;
Matches 361; Conservative 81; Mismatches 145; Indels 133; Gaps 20;



RESULT 12
TVYUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Mart
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specif

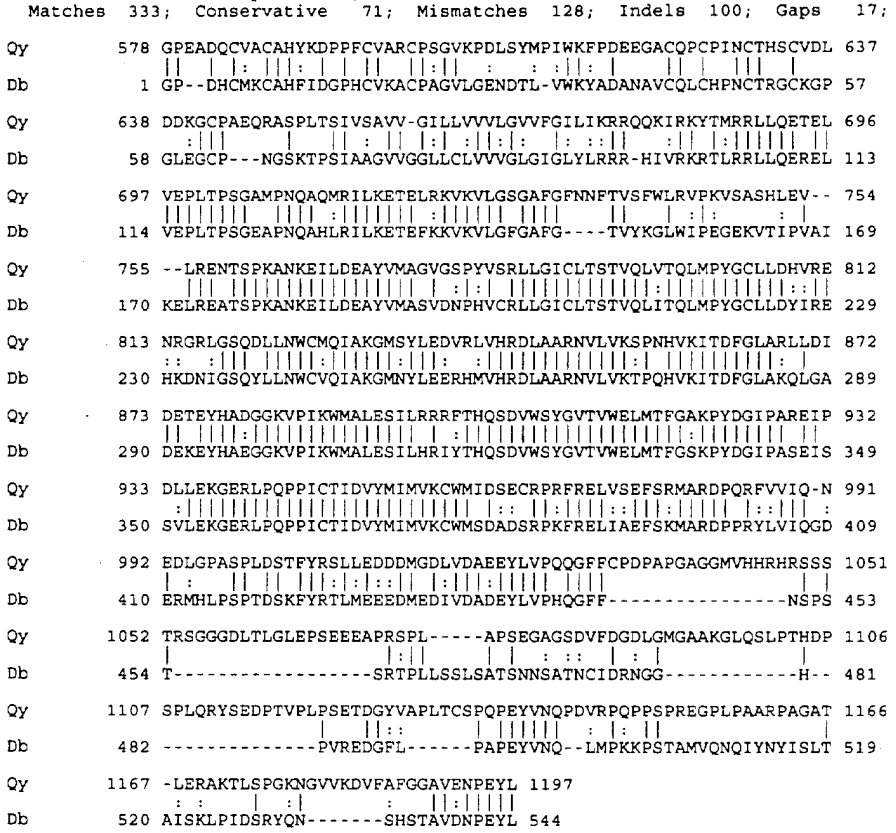
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Query Match 23.4%; Score 1592; DB 1; Length 604;
Best Local Similarity 50.0%; Pred. No. 1e-61;
Matches 347; Conservative 77; Mismatches 136; Indels 134; Gaps 18;



RESULT 13
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
R;Vennstroem, B.
submitted to the EMBL Data Library, March 1993
A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X12707
C;Genetics:
A;Gene: erbB
C;Superfamily:. epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-speci
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 22.6%; Score 1541; DB 2; Length 544;
Best Local Similarity 52.7%; Pred. No. 1.4e-59;
Matches 333; Conservative 71; Mismatches 128; Indels 100; Gaps 17;



RESULT 14
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus m
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

Query Match 22.5%; Score 1534; DB 2; Length 545;
Best Local Similarity 52.7%; Pred. No. 2.8e-59;
Matches 333; Conservative 70; Mismatches 129; Indels 100; Gaps 17;

```
Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
Db          1 GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57

Qy        638 DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETEL 696
Db         58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQEREL 113

Qy        697 VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV-- 754
Db        114 VEPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAFG----TVYKGLWIPEGEKVTIPVAI 169

Qy        755 --LRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRE 812
Db        170 KELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIRE 229

Qy        813 NRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDI 872
Db        230 HKDNIGSQYLLNWCVQIAKGMNYLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQLGA 289

Qy        873 DETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIP 932
Db        290 DEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEIS 349

Qy        933 DLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-N 991
Db        350 SVLEKGERLPQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGD 409

Qy        992 EDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSS 1051
Db        410 ERMHLPSPTDSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPS 453

Qy       1052 TRSGGGDLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDP 1106
Db        454 T-----------------SRTPLLSSLSATSNNSATNCIDRNGG------------H-- 481

Qy       1107 SPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT 1166
Db        482 --------------PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLT 519

Qy       1167 -LERAKTLSPGKNGVVKDVFAFGGAVENPEYL 1197
Db        520 AISKLPMDSRYQN-------SHSTAVDNPEYL 544
```

RESULT 15
B44776
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)
C;Species: avian erythroblastosis virus
C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
C;Accession: B44776
R;Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.
Oncogene 5, 15-24, 1990
A;Title: Six amino acids from the retroviral gene gag greatly enhance the transforming p
A;Reference number: A44776; MUID:90206603; PMID:1969616
A;Accession: B44776
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-540 <BRU>
A;Cross-references: GB:X52211
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif

Query Match 22.3%; Score 1517; DB 2; Length 540;
Best Local Similarity 52.6%; Pred. No. 1.5e-58;
Matches 328; Conservative 70; Mismatches 127; Indels 98; Gaps 16;

```
Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSP 761
Db        118 APNQAHLRILKETEFKKVKVLGFGAFG----TVYKGLWIPEGEKVTIPVAIKELREATSP 173

Qy        762 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 821
Db        174 KANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQY 233

Qy        822 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 881
Db        234 LLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAKQLGADEKEYHAEG 293

Qy        882 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 941
Db        294 GKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERL 353

Qy        942 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPL 1000
Db        354 PQPPICTIDVYMIMVKCWMSGADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPT 413

Qy       1001 DSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLT 1060
Db        414 DSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-------- 449

Qy       1061 LGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSED 1115
Db        450 ----------SRTPLLSSLSATSNNSATNCIDRNGG------------H---------- 476

Qy       1116 PTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERAKTLS 1174
Db        477 -----PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAISKLPMDS 523

Qy       1175 PGKNGVVKDVFAFGGAVENPEYL 1197
Db        524 RYQN-------SHSTAVDNPEYL 539
```

Search completed: July 22, 2003, 09:28:12
Job time : 30.2855 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 Seconds
(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-710-730-14
Perfect score: 6813
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6602	96.9	1255	1	ERB2_HUMAN	P04626 homo sapien
2	5811	85.3	1257	1	ERB2_RAT	P06494 rattus norv
3	5795.5	85.1	1254	1	ERB2_MESAU	Q60553 mesocricetu
4	3033	44.5	1210	1	EGFR_HUMAN	P00533 homo sapien
5	3007	44.1	1210	1	EGFR_MOUSE	Q01279 mus musculu
6	2852.5	41.9	1308	1	ERB4_HUMAN	Q15303 homo sapien
7	2833	41.6	1308	1	ERB4_RAT	Q62956 rattus norv
8	2575.5	37.8	1167	1	XMRK_XIPMA	P13388 xiphophorus
9	2316.5	34.0	1342	1	ERB3_HUMAN	P21860 homo sapien
10	2244.5	32.9	1339	1	ERB3_RAT	Q62799 rattus norv
11	1836	26.9	1426	1	EGFR_DROME	P04412 drosophila
12	1638.5	24.0	634	1	ERBB_ALV	P00534 avian leuko
13	1595	23.4	703	1	EGFR_CHICK	P13387 gallus gall
14	1592	23.4	604	1	ERBB_AVIER	P00535 avian eryth
15	1524	22.4	540	1	ERBB_AVIEU	P11273 avian eryth
16	1217	17.9	1323	1	LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1	ERB2_MOUSE	P70424 mus musculu
18	693	10.2	1363	1	ILPR_BRALA	O02466 branchiosto
19	656.5	9.6	1372	1	INSR_MOUSE	P15208 mus musculu
20	656	9.6	1300	1	IRR_MOUSE	Q9wtl4 mus musculu
21	655	9.6	1382	1	INSR_HUMAN	P06213 homo sapien
22	653	9.6	1383	1	INSR_RAT	P15127 rattus norv
23	647	9.5	1297	1	IRR_HUMAN	P14616 homo sapien
24	644.5	9.5	1300	1	IRR_CAVPO	P14617 cavia porce
25	632	9.3	1477	1	HTK7_HYDAT	Q25197 hydra atten
26	630	9.2	1607	1	MIPR_LYMST	Q25410 lymnaea sta
27	602	8.8	1367	1	IG1R_HUMAN	P08069 homo sapien
28	584	8.6	1373	1	IG1R_MOUSE	Q60751 mus musculu
29	580.5	8.5	1370	1	IG1R_RAT	P24062 rattus norv
30	580	8.5	1114	1	RET_HUMAN	P07949 homo sapien
31	579	8.5	987	1	EPB4_HUMAN	P54760 homo sapien
32	575	8.4	2146	1	INSR_DROME	P09208 drosophila
33	569	8.4	1390	1	INSR_AEDAE	Q93105 aedes aegyp
34	559	8.2	987	1	EPB4_MOUSE	P54761 mus musculu
35	551.5	8.1	984	1	EPB1_CHICK	Q07494 gallus gall
36	548.5	8.1	984	1	EPB1_RAT	P09759 rattus norv
37	544	8.0	902	1	EPBB_XENLA	Q91736 xenopus lae
38	542.5	8.0	984	1	EPB1_HUMAN	P54762 homo sapien
39	536.5	7.9	985	1	EPBA_XENLA	Q91571 xenopus lae
40	534.5	7.8	977	1	EPA2_MOUSE	Q03145 mus musculu
41	529.5	7.8	976	1	EPA2_HUMAN	P29317 homo sapien
42	525.5	7.7	985	1	EP4B_XENLA	Q91694 xenopus lae
43	525.5	7.7	1053	1	FAK1_CHICK	Q00944 gallus gall
44	524.5	7.7	986	1	EP4A_XENLA	Q91845 xenopus lae
45	523	7.7	1068	1	FAK1_XENLA	Q91738 xenopus lae

ALIGNMENTS

RESULT 1
ERB2_HUMAN
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

```
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
CC       POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC       OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC       ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M11767; AAA35808.1; -.
DR   EMBL; M11761; AAA35808.1; JOINED.
DR   EMBL; M11762; AAA35808.1; JOINED.
DR   EMBL; M11763; AAA35808.1; JOINED.
DR   EMBL; M11764; AAA35808.1; JOINED.
DR   EMBL; M11765; AAA35808.1; JOINED.
DR   EMBL; M11766; AAA35808.1; JOINED.
DR   EMBL; M11730; AAA75493.1; -.
DR   EMBL; M12036; AAA35978.1; -.
DR   EMBL; X03363; CAA27060.1; -.
DR   PIR; A25491; A25491.
DR   PIR; A24571; A24571.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3430; ERBB2.
DR   MIM; 164870; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Polymorphism.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1255       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    220    227       BY SIMILARITY.
FT   DISULFID    224    235       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     654    654       I -> V.
FT                                /FTId=VAR_004077.
FT   VARIANT     655    655       I -> V.
FT                                /FTId=VAR_004078.
FT   CONFLICT   1170   1170       P -> A (IN REF. 2).
SQ   SEQUENCE   1255 AA;  137909 MW;  39E9DFDA04DCF962 CRC64;

  Query Match             96.9%;  Score 6602;  DB 1;  Length 1255;
  Best Local Similarity   96.9%;  Pred. No. 0;
  Matches 1221;  Conservative   11;  Mismatches   18;  Indels   10;  Gaps    3;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||||||||||||||||||||||||||||||||||||||||:::::
Db        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
```

```
Qy        721 RKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANKEILDEAYVMA 775
              |||||||||||||    ||   : :|     |:  : ::|||||||||||||||||||||
Db        721 RKVKVLGSGAFG----TVYKGIWIPDGENVKIPVA-IKVLRENTSPKANKEILDEAYVMA 775

Qy        776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835

Qy        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895

Qy        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955

Qy        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1015

Qy       1016 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1075

Qy       1076 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135

Qy       1136 QPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPE 1195
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1136 QPEYVNQPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPE 1195

Qy       1196 YLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 YLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
```

```
RESULT 2
ERB2_RAT
ID   ERB2_RAT       STANDARD;      PRT;  1257 AA.
AC   P06494;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE   receptor-related protein).
GN   ERBB2 OR NEU.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=86118662; PubMed=3945311;
RA   Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT   "The neu oncogene encodes an epidermal growth factor receptor-related
RT   protein.";
RL   Nature 319:226-230(1986).
RN   [2]
RP   SEQUENCE OF 852-905 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   STRUCTURE BY NMR OF 650-668.
RX   MEDLINE=92155181; PubMed=1346763;
RA   Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,
RA   Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT   "Three dimensional structure of the transmembrane region of the proto-
RT   oncogenic and oncogenic forms of the neu protein.";
RL   EMBO J. 11:43-48(1992).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC       THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03362; CAA27059.1; ALT_INIT.
DR   PIR; A24562; TVRTNU.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1257       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    677       POTENTIAL.
FT   DOMAIN      678   1257       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      159    369       CYS-RICH.
FT   DOMAIN      473    646       CYS-RICH.
FT   DOMAIN      722    989       PROTEIN KINASE.
FT   NP_BIND     728    736       ATP (BY SIMILARITY).
FT   BINDING     755    755       ATP (BY SIMILARITY).
FT   ACT_SITE    847    847       BY SIMILARITY.
FT   DISULFID    196    205       BY SIMILARITY.
FT   DISULFID    200    213       BY SIMILARITY.
FT   DISULFID    221    228       BY SIMILARITY.
FT   DISULFID    225    236       BY SIMILARITY.
FT   DISULFID    237    245       BY SIMILARITY.
FT   DISULFID    241    253       BY SIMILARITY.
FT   DISULFID    256    265       BY SIMILARITY.
FT   DISULFID    269    296       BY SIMILARITY.
FT   DISULFID    300    312       BY SIMILARITY.
FT   DISULFID    316    332       BY SIMILARITY.
FT   DISULFID    335    339       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    578       BY SIMILARITY.
```

```
FT   DISULFID    569    586       BY SIMILARITY.
FT   DISULFID    589    598       BY SIMILARITY.
FT   DISULFID    602    625       BY SIMILARITY.
FT   DISULFID    628    636       BY SIMILARITY.
FT   DISULFID    632    644       BY SIMILARITY.
FT   MOD_RES    1141   1141       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1250   1250       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    260    260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     661    661       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1257 AA;  138831 MW;  6129264583011402 CRC64;

  Query Match             85.3%;  Score 5811;  DB 1;  Length 1257;
  Best Local Similarity   85.3%;  Pred. No. 6.8e-304;
  Matches 1077;  Conservative   58;  Mismatches  115;  Indels   12;  Gaps    5;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||| |||||||| | ||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGFLLALLPPGIAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              ||||:| |||||||||||||||:|||||||::|||||||||||||||||| ||||||||
Db         61 ELTYVPANASLSFLQDIQEVQGYMLIAHNQVKRVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVT-GASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
              || :|    | | :| |||||||||||||||||||: |||||||| :||||:| |||||
Db        121 DPQDNVAASTPGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDMVLWKDVFRKNNQL 180

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQ 239
              |   |||||||| ||:| || : ||||| |||| || |:|  ||||||| |||||||||
Db        181 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCCHEQ 240

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||| |
Db        241 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMHNPEGRYTFGASCVTTC 300

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |||||||:|||||||| :|||||||||||||||||||||||||||:::::
Db        301 PYNYLSTEVGSCTLVCPPNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLRGARAITSD 360

Qy        360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 419
                || |||||||||||||||||||:|  |||:|||||||||||||||||||||||||||
Db        361 NVQEFDGCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 420

Qy        420 PDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHT 479
              ||||||||::|||||||:|||||||||||||  |||||||||||||||| | |||||||
Db        421 RDLSVFQNLRIIRGRILHDGAYSLTLQGLGIHSLGLRSLRELGSGLALIHRNAHLCFVHT 480

Qy        480 VPWDQLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |||||||||||||||: |||||:: ||  || |: ||| ||||||||||||||| |||||
Db        481 VPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCWGPGPTQCVNCSHFLRGQ 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||||||| :|||||||: : |||||||||||||| | |||| ||||| |||||||    |||
Db        541 ECVEECRVWKGLPREYVSDKRCLPCHPECQPQNSSETCFGSEADQCAACAHYKDSSSCVA 600

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAV 658
              |||||||||||||||||:|||||| |||||||||||||||::||||||||:| |:: |
Db        601 RCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQRASPVTFIIATV 660

Qy        659 VGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 718
              ||:|| ::| || |||||||:|||||||||||||||||||||||||||||||||||||||
Db        661 VGVLLFLILVVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKET 720

Qy        719 ELRKVKVLGSGAFGTVNKGVWIPEGETVKIPVAIKELREATSPKANKEILDEAYV 773
              |||||||||||||    ||   : :|     |:  : ::|||||||||||||||||||
Db        721 ELRKVKVLGSGAFG----TVYKGIWIPDGENVKIPVA-IKVLRENTSPKANKEILDEAYV 775

Qy        774 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGM 833
              ||||||||||||||||||||||||||||||||||||||:|||||||||||:||||||||
Db        776 MAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGM 835

Qy        834 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 SYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALES 895

Qy        894 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 953
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 ILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYM 955

Qy        954 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDD 1013
              ||||||||||||||||||||||||||||||||||||||||:||:|||||||||||||
Db        956 IMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDDD 1015

Qy       1014 MGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS 1073
              |||||||||||||||||| ||| || |   | ||||||||||:|||||||||  |||
Db       1016 MGDLVDAEEYLVPQQGFFSPDPTPGTGSTAHRRHRSSSTRSGGGELTLGLEPSEEGPPRS 1075

Qy       1074 PLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1133
              |||||||||||||||| ||  ||||||  || ||||||||||:||| |||||||| |
Db       1076 PLAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDLSPLQRYSEDPTLPLPPETDGYVAPLAC 1135

Qy       1134 SPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVEN 1193
              |||||||| :|:|||| |||||  |||||||||| ||||||||||||||||||||||||
Db       1136 SPQPEYVNQSEVQPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFAFGGAVEN 1195

Qy       1194 PEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1253
              |||| |: | |  ||| |||||||||||||:  |:| ||| |:|||||||||||||||
Db       1196 PEYLVPREGTASPPHPSPAFSPAFDNLYYWDQNSSEQGPPPSNFEGTPTAENPEYLGLDV 1255

Qy       1254 PV 1255
              ||
Db       1256 PV 1257


RESULT 3
ERB2_MESAU
ID   ERB2_MESAU     STANDARD;      PRT;  1254 AA.
AC   Q60553;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN   ERBB2 OR NEU.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Nerve;
RX   MEDLINE=94193007; PubMed=7908275;
RA   Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA   Yamazaki Y., Ishikawa T.;
RT   "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL   Gene 140:251-255(1994).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC       ALTHOUGHT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC       POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC       ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D16295; BAA03801.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Proto-oncogene; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1254       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    653    675       POTENTIAL.
FT   DOMAIN      676   1254       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      158    368       CYS-RICH.
FT   DOMAIN      472    644       CYS-RICH.
FT   DOMAIN      720    987       PROTEIN KINASE.
FT   NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1247   1247       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     658    658       V -> E (IN ONCOGENIC NEU).
FT   VARIANT     659    659       V -> E (IN ONCOGENIC NEU).
SQ   SEQUENCE   1254 AA;  138252 MW;  974C3791C21F2BE1 CRC64;

  Query Match             85.1%;  Score 5795.5;  DB 1;  Length 1254;
  Best Local Similarity   84.9%;  Pred. No. 4.6e-303;
  Matches 1070;  Conservative   67;  Mismatches  112;  Indels   11;  Gaps    4;
```

```
Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| | |||||||| |||: |||||||||||||||||||||::||||||||||||||
Db          1 MELAAWCGWGLLLALLSPGASGTQVCTGTDMKLRLPASPETHLDIVRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||:|||||||||||||:||||:||| |||||||||||||||| ||||||||
Db         61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTQLFEDKYALAVLDNR 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||:| |  || :| ||||||||||||||||||: |||||||||:||||:| ||||||
Db        121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIRGNPQLCYQDTVLWKDVFRKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
                |||||||| ||:| || : ||| | ||||:|| |:         :  ||||||||||
Db        181 PVDIDTNRSRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              ||||||:|||||||||:|||||||||||||||||| ||||||||||::::
Db        301 YNYLSTEVGSCTLVCPLNNQEVTAEDGTQRCEKCSKSCARVCYGLGMEHLRGARAITSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
                |||||||||||||||||||:|:|  ||| ||||||||||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGNPSSGIAPLTPEQLQVFETLEEITGYLYISAWPDSLH 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              ||||||||:|||||:||:|||||  |||||| ||||||||||||| ||| ||||||||||
Db        421 DLSVFQNLRVIRGRVLHDGAYSLALQGLGIRWLGLRSLRELGSGLVLIHRNTHLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||||: |  |:||  :  ||: ||| |||||||||||||| |||||||
Db        481 PWDQLFRNPHQALLHSGNPSEEECGLKDFACYPLCAHGHCWGPGPTQCVNCSHFLRGQEC 540

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |:|||| :|||||||| :|||||||||||||| : || | ||||| || |||| |||||||
Db        541 VKECRVWKGLPREYVNGKHCLPCHPECQPQNSTETCTGSEADQCTACPHYKDSPFCVARC 600

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              ||||||||||||||:||||| |||||||||||||||::|||||||||| |||:: |||
Db        601 PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIIATVVG 660

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              ||| :|:||| |||||||:|||||||||||||||||||||||||||||||||||||||||
Db        661 ILLFLVIGVVVGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy        721 RKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANKEILDEAYVMA 775
              |||||||||||     ||   : :|     |:  : ::|||||||||||||||||||||
Db        721 RKVKVLGSGAFG----TVYKGIWIPDGENVKIPVA-IKVLRENTSPKANKEILDEAYVMA 775

Qy        776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835
              |:|||||||||||||||||||||||||||||||:|||||||||||||:||||||||
Db        776 GLGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSY 835

Qy        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWIALESIL 895

Qy        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955

Qy        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1015
              |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDMG 1015

Qy       1016 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1075
              |||||||||||||| ||||||||   | ||||||||||:||||:||| || |||||
Db       1016 DLVDAEEYLVPQQGFFFPDPAPGAGSTAHRRHRSSSTRSGGGELTLGMEPSGEEPPRSPL 1075
```

```
Qy       1076 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
              ||||||||||:|:|||| || ||:   | ||||||||||:|||:|||||||| |||
Db       1076 APSEGAGSDVFEGELGMGATKGPQSISPRDLSPLQRYSEDPTLPLPTETDGYVAPLACSP 1135

Qy       1136 QPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPE 1195
              ||||||||:||||||   |||||  |||||||| ||||||||||||||| |||||||||
Db       1136 QPEYVNQPEVRPQPPLTPEGPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAVENPE 1195

Qy       1196 YLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
              || |:||:| ||| |||  ||||||||||||| |||:||:||:|||||||||||||||||
Db       1196 YLVPRGGSASQPH-PPALCPAFDNLYYWDQDPSERGSPPNTFEGTPTAENPEYLGLDVPV 1254


RESULT 4
EGFR_HUMAN
ID   EGFR_HUMAN     STANDARD;      PRT;  1210 AA.
AC   P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC   O00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE   protein-tyrosine kinase ErbB-1).
GN   EGFR OR ERBB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=84219729; PubMed=6328312;
RA   Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA   Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA   Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT   "Human epidermal growth factor receptor cDNA sequence and aberrant
RT   expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL   Nature 309:418-425(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=95382957; PubMed=7654368;
RA   Ilekis J.V., Stark B.C., Scoccia B.;
RT   "Possible role of variant RNA transcripts in the regulation of
RT   epidermal growth factor receptor expression in human placenta.";
RL   Mol. Reprod. Dev. 41:149-156(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97078686; PubMed=8918811;
RA   Reiter J.L., Maihle N.J.;
RT   "A 1.8 kb alternative transcript from the human epidermal growth
RT   factor receptor gene encodes a truncated form of the receptor.";
RL   Nucleic Acids Res. 24:4050-4056(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97256547; PubMed=9103388;
RA   Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT   "Expression of a truncated epidermal growth factor receptor-like
RT   protein (TEGFR) in ovarian cancer.";
RL   Gynecol. Oncol. 65:36-41(1997).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC   TISSUE=Placenta;
RX   MEDLINE=21100872; PubMed=11161793;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative EGFR transcripts encoding truncated receptor
RT   isoforms.";
RL   Genomics 71:1-20(2001).
RN   [6]
RP   SEQUENCE OF 575-687 FROM N.A.
RA   Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA   Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA   Maihle N.J.;
RT   "Human and mouse alternative EGFR transcripts encoding only the
RT   extracellular domain of the receptor.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 713-924 FROM N.A.
RX   MEDLINE=84196372; PubMed=6326261;
RA   Lin C.R., Chen W.S., Kruiger W., Stolarsky L.S., Weber W.,
RA   Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT   "Expression cloning of human EGF receptor complementary DNA: gene
RT   amplification and three related messenger RNA products in A431
RT   cells.";
RL   Science 224:843-848(1984).
RN   [8]
RP   SEQUENCE OF 150-962 FROM N.A.
RX   MEDLINE=84245835; PubMed=6330563;
RA   Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA   Roe B.A., Merlino G.T., Pastan I.;
RT   "Human epidermal growth factor receptor cDNA is homologous to a
RT   variety of RNAs overproduced in A431 carcinoma cells.";
RL   Nature 309:806-810(1984).
RN   [9]
RP   SEQUENCE OF 1028-1210 FROM N.A.
RX   MEDLINE=85046483; PubMed=6093780;
RA   Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA   O'Malley B.W.;
RT   "Isolation of an evolutionarily conserved epidermal growth factor
RT   receptor cDNA from human A431 carcinoma cells.";
RL   Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN   [10]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=88217333; PubMed=3329716;
RA   Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA   Waterfield M.D.;
RT   "The human EGF receptor gene: structure of the 110 kb locus and
RT   identification of sequences regulating its transcription.";
RL   Oncogene Res. 1:375-396(1987).
RN   [11]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=91107677; PubMed=1988448;
RA   Haley J.D., Waterfield M.D.;
RT   "Contributory effects of de novo transcription and premature
RT   transcript termination in the regulation of human epidermal growth
RT   factor receptor proto-oncogene RNA synthesis.";
RL   J. Biol. Chem. 266:1746-1753(1991).
RN   [12]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=85270438; PubMed=2991899;
RA   Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT   "Characterization and sequence of the promoter region of the human
RT   epidermal growth factor receptor gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN   [13]
RP   SEQUENCE OF 540.
RA   Kohda D.;
RL   Submitted (SEP-1997) to the SWISS-PROT data bank.
RN   [14]
RP   RECEPTOR ACTIVITY.
RX   MEDLINE=84191554; PubMed=6325948;
RA   Mroczkowski B., Mosig G., Cohen S.;
RT   "ATP-stimulated interaction between epidermal growth factor receptor
RT   and supercoiled DNA.";
RL   Nature 309:270-273(1984).
RN   [15]
RP   PHOSPHORYLATION.
RX   MEDLINE=89278137; PubMed=2543678;
RA   Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA   Howk R., Givol D., Ullrich A., Schlessinger J.;
RT   "All autophosphorylation sites of epidermal growth factor (EGF)
```

```
RT   receptor and HER2/neu are located in their carboxyl-terminal tails.
RT   Identification of a novel site in EGF receptor.";
RL   J. Biol. Chem. 264:10667-10671(1989).
RN   [16]
RP   CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
RP   ASN-528.
RX   MEDLINE=96398132; PubMed=8962717;
RA   Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
RT   "Analysis of the glycosylation patterns of the extracellular domain of
RT   the epidermal growth factor receptor expressed in Chinese hamster
RT   ovary fibroblasts.";
RL   Growth Factors 13:121-132(1996).
RN   [17]
RP   CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
RP   ASN-603.
RX   MEDLINE=20198209; PubMed=10731668;
RA   Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
RT   "Characterization of the N-oligosaccharides attached to the atypical
RT   Asn-X-Cys sequence of recombinant human epidermal growth factor
RT   receptor.";
RL   J. Biochem. 127:65-72(2000).
RN   [18]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=98225196; PubMed=9556602;
RA   Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
RT   "Disulfide bond structure of human epidermal growth factor receptor.";
RL   J. Biol. Chem. 273:11150-11157(1998).
RN   [19]
RP   REVIEW.
RX   MEDLINE=87297456; PubMed=3039909;
RA   Carpenter G.;
RT   "Receptors for epidermal growth factor and other polypeptide
RT   mitogens.";
RL   Annu. Rev. Biochem. 56:881-914(1987).
CC   -!- FUNCTION: Receptor for EGF, but also for other members of the EGF
CC       family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
CC       EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
CC       involved in the control of cell growth and differentiation.
CC   -!- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
CC       secreted.
CC   -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
CC       truncated isoform/TEGFR, 3/p110 and 4; are produced by
CC       alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
CC       expressed in ovarian cancers.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00588; CAA25240.1; -.
DR   EMBL; U95089; AAB53063.1; -.
DR   EMBL; U48722; AAC50802.1; -.
DR   EMBL; U48723; AAC50804.1; -.
DR   EMBL; U48724; AAC50796.1; -.
DR   EMBL; U48725; AAC50797.1; -.
DR   EMBL; U48726; AAC50798.1; -.

  Query Match             44.5%;  Score 3033;  DB 1;  Length 1210;
  Best Local Similarity   48.7%;  Pred. No. 3.7e-155;
  Matches  619;  Conservative  175;  Mismatches  360;  Indels  116;  Gaps   25;
```

```
Qy      11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db      14 LLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN 73

Qy      69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db      74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALAVLSNYD------- 126

Qy     129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNR 188
Db     127 ---ANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183

Qy     189 SRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTGP 247
Db     184 LGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGP 243

Qy     248 KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
Db     244 RESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVTD 303

Qy     308 VGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
Db     304 HGSCVRACGADSYEM-EEDGVRKCKKCEGPCRKVCNGIGIGEFK-DSLSINATNIKHFKN 361

Qy     367 CKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLSVFQ 426
Db     362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFE 421

Qy     427 NLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLF 486
Db     422 NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKLF 481

Qy     487 RNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRV 546
Db     482 GTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNL 541

Qy     547 LQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPSGVKP 606
Db     542 LEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMG 601

Qy     607 DLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG---ILL 663
Db     602 ENNTL-VWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIP--SIATGMVGALLLL 658

Qy     664 VVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKV 723
Db     659 VVALGIG---LFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKI 715

Qy     724 KVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSPKANKEILDEAYVMAGVGS 779
Db     716 KVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDN 771

Qy     780 PYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDV 839
Db     772 PHVCRLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDR 831

Qy     840 RLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRF 899
Db     832 RLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIY 891

Qy     900 THQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCW 959
Db     892 THQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCW 951

Qy     960 MIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLV 1018
Db     952 MIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVV 1011

Qy    1019 DAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPS 1078
Db    1012 DADEYLIPQQGFF----------------------------------SSPSTSRTPLLSS 1037
```

```
Qy       1079 EGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQ 1136
              | |:  :  :      ||||  |  : | ||||| |||  |  ::  | ::      |
Db       1038 LSATSN--NSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL------PV 1089

Qy       1137 PEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY 1196
              |||:||  |  :|    : |:   :|           :| ::   :|      || ||||
Db       1090 PEYINQ-SVPKRPAGSVQNPVYHNQPLNP--------APSRDPHYQD--PHSTAVGNPEY 1138

Qy       1197 L-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTFKG 1239
              | | |      |   :  ||:  :| |          :|       |:   |   |||
Db       1139 LNTVQ---------PTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKG 1189

Qy       1240 TPTAENPEYL 1249
              : |||| |||
Db       1190 S-TAENAEYL 1198
```

RESULT 5
EGFR_MOUSE
ID   EGFR_MOUSE   STANDARD;   PRT; 1210 AA.
AC   Q01279;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=93026370; PubMed=1408137;
RA   Avivi A., Skorecki K., Yayon A., Givol D.;
RT   "Promoter region of the murine fibroblast growth factor receptor 2
RT   (bek/KGFR) gene.";
RL   Oncogene 7:1957-1962(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX   MEDLINE=93126380; PubMed=7678348;
RA   Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT   "Expression of the epidermal growth factor receptor gene is regulated
RT   in mouse blastocysts during delayed implantation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RA   Hibbs M.L.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/C3; TISSUE=Liver;
RX   MEDLINE=94170986; PubMed=8125255;
RA   Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA   Jenkins N.A., Lee D.C.;
RT   "The mouse waved-2 phenotype results from a point mutation in the EGF
RT   receptor tyrosine kinase.";
RL   Genes Dev. 8:399-413(1994).
RN   [5]
RP   SEQUENCE OF 1-714 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=91232866; PubMed=2030916;
RA   Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT   "Comparison of EGF receptor sequences as a guide to study the ligand
RT   binding site.";
RL   Oncogene 6:673-676(1991).
RN   [6]
RP   SEQUENCE OF 969-1117 FROM N.A.
RC   STRAIN=C3H;
RA   Eisinger D.P., Serrero G.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.

CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC   AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC   VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC   tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC   dimerization, internalization of the EGF-receptor complex,
CC   induction of the tyrosine kinase activity, stimulation of cell DNA
CC   synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X78987; CAA55587.1; -.
DR   EMBL; U03425; AAA17899.1; -.
DR   EMBL; X59698; CAA42219.1; -.
DR   EMBL; L06864; AAA53029.1; -.
DR   EMBL; Z12608; CAA78249.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT   SIGNAL   1   24   POTENTIAL.
FT   CHAIN   25   1210   EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN   25   647   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   648   670   POTENTIAL.
FT   DOMAIN   671   1210   CYTOPLASMIC (POTENTIAL).
FT   REPEAT   75   300   APPROXIMATE.
FT   REPEAT   390   600   APPROXIMATE.
FT   DOMAIN   1028   1071   SER-RICH.
FT   DOMAIN   714   981   PROTEIN KINASE.
FT   NP_BIND   720   728   ATP (BY SIMILARITY).
FT   BINDING   747   747   ATP (BY SIMILARITY).
FT   ACT_SITE   839   839   BY SIMILARITY.
FT   DISULFID   190   199   BY SIMILARITY.
FT   DISULFID   194   207   BY SIMILARITY.
FT   DISULFID   215   223   BY SIMILARITY.
FT   DISULFID   219   231   BY SIMILARITY.
FT   DISULFID   232   240   BY SIMILARITY.
FT   DISULFID   236   248   BY SIMILARITY.
FT   DISULFID   251   260   BY SIMILARITY.
FT   DISULFID   264   291   BY SIMILARITY.
FT   DISULFID   295   307   BY SIMILARITY.
FT   DISULFID   311   326   BY SIMILARITY.
FT   DISULFID   329   333   BY SIMILARITY.
FT   DISULFID   506   515   BY SIMILARITY.
FT   DISULFID   510   523   BY SIMILARITY.
FT   DISULFID   526   535   BY SIMILARITY.
FT   DISULFID   539   555   BY SIMILARITY.
FT   DISULFID   558   571   BY SIMILARITY.
FT   DISULFID   562   579   BY SIMILARITY.
FT   DISULFID   582   591   BY SIMILARITY.
FT   DISULFID   595   617   BY SIMILARITY.

```
FT   DISULFID    620    628       BY SIMILARITY.
FT   DISULFID    624    636       BY SIMILARITY.
FT   MOD_RES     680    680       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   MOD_RES    1092   1092       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1110   1110       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1172   1172       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1197   1197       PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT                                (BY SIMILARITY).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    175    175       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    352    352       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    444    444       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     19     19       C -> S (IN REF. 2).
FT   CONFLICT    539    539       C -> W (IN REF. 5).
FT   CONFLICT    991    991       L -> F (IN REF. 4).
FT   CONFLICT   1116   1117       HP -> DR (IN REF. 6).
SQ   SEQUENCE   1210 AA;  134853 MW;  690E20D46DF2D2F5 CRC64;

  Query Match             44.1%;  Score 3007;  DB 1;  Length 1210;
  Best Local Similarity   48.6%;  Pred. No. 9e-154;
  Matches  621;  Conservative  169;  Mismatches  367;  Indels  120;  Gaps   27;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
              ||  ||    |  |     :|| ||  :|    : | |   |: :|  |:|| ||||:||:  |
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
                ||||: |||| |||||| | | ::||: |:|:||  |:|: ||||:| |
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
               |  :  ||||| :|:| ||| | |    || ||  ||| |:||    |  |  : |
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
               :  :|  | | |    |||   |:|| ||: :||  |: ||:|  |:|||| |||||
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              ||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:||  || ||
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              : || |||  |   ||  ||| ::|:||  || :|| |:|:  | ::  |  | ::
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
               |  |  | | |  || :| ||  :  | || | :|:: :|::||||:| | ||||:  ||
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
               |:||::||||   :| :|| : || |: ||||||:|:  |  :|  | :||: :|: |
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
              :||  |:|      || | :|      |: ||:   |||| |  ||:|    ||:||||
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
              :| :|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||  ||:
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
              |:  : : : :||: |    |  | |||: |     :||       : || : :|| |
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 LVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELR 721
```

```
                | :|: |  || |  ||:  :|| |:|||||| |||||||||  |||| :|||||| :
Db        657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFK 715

Qy        722 KVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSPKANKEILDEAYVMAGV 777
              |:|||||||||   ||  | :|:    : |   ||| |||||||||||||||||| |
Db        716 KIKVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASV 771

Qy        778 GSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLE 837
               :|:| |||||||||||||:|||||||||||:|||::  :||| |||||:||||||:|||
Db        772 DNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLE 831

Qy        838 DVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRR 897
              | |||||||||||||||:| ||||||||||:||  :| ||||:|||||||||||||| |
Db        832 DRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHR 891

Qy        898 RFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVK 957
              :|||||||||||||||||||:|||||||| :|  :|||||||||||||||||||||||||
Db        892 IYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVK 951

Qy        958 CWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGD 1016
              |||||:: ||:||||: |||:|||||||::||| :| :   || || |||:|::::|| |
Db        952 CWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMED 1011

Qy       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076
              :|||:|||:|||||                | ||                  |:||
Db       1012 VVDADEYLIPQQGFF----------------NSPST-----------------SRTPLL 1037

Qy       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCS 1134
               |  | |:    :  :       |    : : ||||| |||  :  :   | ::
Db       1038 SSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------ 1087

Qy       1135 PQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENP 1194
              | ||||||  |  :|    : |:   :|           :||::   ::      || ||
Db       1088 PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNP 1136

Qy       1195 EYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTF 1237
              ||| | |         |  |  |::   | |          :|       |:  |  |
Db       1137 EYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIF 1187

Qy       1238 KGTPTAENPEYLGLDVP 1254
              || ||||| ||| :  |
Db       1188 KG-PTAENAEYLRVAPP 1203
```

```
RESULT 6
ERB4_HUMAN
ID   ERB4_HUMAN     STANDARD;      PRT;  1308 AA.
AC   Q15303;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE   (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN   ERBB4 OR HER4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM JM-A).
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=93189574; PubMed=8383326;
RA   Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA   Foy L., Neubauer M.G., Shoyab M.;
RT   "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT   epidermal growth factor receptor family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97476287; PubMed=9334263;
RA   Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA   Klagsbrun M.;
```

```
RT   "A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
RT   tissue distribution and differential processing in response to
RT   phorbol ester.";
RL   J. Biol. Chem. 272:26761-26768(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC       FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC       BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC       NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC       CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC       KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC       PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC       LUNG, SALIVARY GLAND, AND PANCREAS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07868; AAB59446.1; -.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3432; ERBB4.
DR   MIM; 600543; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL       1    25       POTENTIAL.
FT   CHAIN       26  1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN      26   651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   652   675       POTENTIAL.
FT   DOMAIN     676  1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     186   334       CYS-RICH.
FT   DOMAIN     496   633       CYS-RICH.
FT   DOMAIN     718   985       PROTEIN KINASE.
FT   NP_BIND    724   732       ATP (BY SIMILARITY).
FT   BINDING    751   751       ATP (BY SIMILARITY).
FT   ACT_SITE   843   843       BY SIMILARITY.
FT   DISULFID   189   197       BY SIMILARITY.
FT   DISULFID   193   205       BY SIMILARITY.
FT   DISULFID   213   221       BY SIMILARITY.
FT   DISULFID   217   229       BY SIMILARITY.
FT   DISULFID   230   238       BY SIMILARITY.
FT   DISULFID   234   246       BY SIMILARITY.
FT   DISULFID   249   258       BY SIMILARITY.
FT   DISULFID   262   289       BY SIMILARITY.
FT   DISULFID   293   304       BY SIMILARITY.
FT   DISULFID   308   323       BY SIMILARITY.
FT   DISULFID   326   330       BY SIMILARITY.
FT   DISULFID   503   512       BY SIMILARITY.
FT   DISULFID   507   520       BY SIMILARITY.
FT   DISULFID   523   532       BY SIMILARITY.
FT   DISULFID   536   552       BY SIMILARITY.
FT   DISULFID   555   569       BY SIMILARITY.
FT   DISULFID   559   577       BY SIMILARITY.
FT   DISULFID   580   589       BY SIMILARITY.
FT   DISULFID   593   614       BY SIMILARITY.
FT   DISULFID   617   625       BY SIMILARITY.
FT   DISULFID   621   633       BY SIMILARITY.
FT   MOD_RES   1162  1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1188  1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1258  1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1284  1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD   138   138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   174   174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   181   181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   253   253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   358   358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   410   410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   473   473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   495   495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   548   548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   576   576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   620   620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   626   648       NGPTSHDCIYYPWTGHSTLPQHA -> IGSSIEDCIGLMD
FT                              (IN ISOFORM JM-B).
SQ   SEQUENCE   1308 AA;  146807 MW;  5E4AE80985D88761 CRC64;
```

```
  Query Match             41.9%;  Score 2852.5;  DB 1;  Length 1308;
  Best Local Similarity   44.1%;  Pred. No. 1.8e-145;
  Matches  596;  Conservative  187;  Mismatches  385;  Indels  183;  Gaps   31;

Qy          9 WGLLLALLPPGAA----STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTY 64
              |  :  |: |       |  || ||: ||   :  |      ||  |: |:|| ||||:|
Db          8 WVWVSLLVAAGTVQPSDSQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVMGNLEITS 67

Qy         65 LPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLN 124
              : |  ||||: ::|| ||||:| || | :||: |||:|||:|:|| ||||:  |
Db         68 IEHNRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIFLNYRKDG 127

Qy        125 NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLI 184
              |          ||:|| |::||||| ||| : :|  ||| ||| |:||        |||:
Db        128 NF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTLV 178

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAG 243
              || |  |  |    | | |||| :   ||:|||||||  |  || ||  :|||| :|| |
Db        179 STNGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG 237

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
              |:||| :|| ||::|| || |   ||    ||  ||:  |  :||:|| ||  ||:|:
Db        238 CSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCVKKCPHNF 297

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
              :  |  ||   ||   ||  |:| : |: |:  | : | |:|   : :          :
Db        298 V-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDK 355

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||     ||| :    : ||:| || |: |||||:| | :|| :: | |
Db        356 FINCTKINGNLIFLVTGIHGDPYNAIEAIDPEKLNVFRTVREITGFLNIQSWPPNMTDFS 415

Qy        424 VFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWD 483
              || ||  | ||:|::|   | |:  ||: |  :|||:|: :|   |  |::||: ||: |
Db        416 VFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCYYHTINWT 475

Qy        484 QLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543
              ||   :| ::  ||  : |  ||: |: ||:   ||||||  ||::| :| ||: |:|
```

```
Db        476 TLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPGPDQCLSCRRFSRGRICIES 535

Qy        544 CRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602

Db        536 CNLYDGEFREFENGSICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPNCVEKCPD 595

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PAEQRASPL 651

Db        596 GLQGANSF--IFKYADPDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQHAR-TPL 652

Qy        652 TSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710

Db        653 --IAAGVIGGLFILVIVGLTFAVYVRRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNQA 708

Qy        711 QMRILKETELRKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANK 765

Db        709 QLRILKETELKRVKVLGSGAFG----TVYKGIWVPEGETVKIPVA-IKILNETTGPKANV 763

Qy        766 EILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNW 825

Db        764 EFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLLNW 823

Qy        826 CMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVP 885

Db        824 CVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKMP 883

Qy        886 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPP 945

Db        884 IKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQPP 943

Qy        946 ICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPASPLDSTF 1004

Db        944 ICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLPSPNDSKF 1003

Qy       1005 YRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLE 1064

Db       1004 FQNLLDEEDLEDMMDAEEYLVP-QAFNIPPP------IYTSRARIDSNRS-----EIGHS 1051

Qy       1065 PSEEEAPRS----------------------------PLAP-SEGAGSDVFDGDLGMGAA 1095

Db       1052 PPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDDSCCNGTL 1111

Qy       1096 KGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQ 1148

Db       1112 RKPVAPHVQEDSSTQRYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPVE---- 1167

Qy       1149 PPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPH 1208

Db       1168 -----ENPFVSRR--------------KNGDLQ-------ALDNPEYHNASNG------ 1194

Qy       1209 PPPA---------------------------------FSPAFDNLYYWDQDPPERGA-- 1232

Db       1195 PPKAEDEYVNEPLYLNTFANTLGKAEYLKNNILSMPEKAKKAFDNPDYWNHSLPPRSTLQ 1254

Qy       1233 PPSTFKGTPT-------------AENPEYL 1249

Db       1255 HPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285


RESULT 7
ERB4_RAT
ID   ERB4_RAT       STANDARD;      PRT;  1308 AA.
AC   Q62956; Q9Z2N7;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN   ERBB4 OR TYRO-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=98221155; PubMed=9553078;
RA   Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA   Marchionni M.A., Kelly R.A.;
RT   "Neuregulins promote survival and growth of cardiac myocytes.
RT   Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT   ventricular myocytes.";
RL   J. Biol. Chem. 273:10261-10269(1998).
RN   [2]
RP   SEQUENCE OF 848-901 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   SEQUENCE OF 1031-1198 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC       2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC       NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC       NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC       RECEPTORS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC       NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC       OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC       RETICULAR NUCLEUS OF THE THALLAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC       HEART.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF041838; AAD08899.1; -.
DR   EMBL; U52531; AAC53051.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 4.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL       1     25       POTENTIAL.
```

```
FT   CHAIN        26   1308       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT   DOMAIN       26    651       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    652    675       POTENTIAL.
FT   DOMAIN      676   1308       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      186    334       CYS-RICH.
FT   DOMAIN      496    633       CYS-RICH.
FT   DOMAIN      718    985       PROTEIN KINASE.
FT   NP_BIND     724    732       ATP (BY SIMILARITY).
FT   BINDING     751    751       ATP (BY SIMILARITY).
FT   ACT_SITE    843    843       BY SIMILARITY.
FT   DISULFID    189    197       BY SIMILARITY.
FT   DISULFID    193    205       BY SIMILARITY.
FT   DISULFID    213    221       BY SIMILARITY.
FT   DISULFID    217    229       BY SIMILARITY.
FT   DISULFID    230    238       BY SIMILARITY.
FT   DISULFID    234    246       BY SIMILARITY.
FT   DISULFID    249    258       BY SIMILARITY.
FT   DISULFID    262    289       BY SIMILARITY.
FT   DISULFID    293    304       BY SIMILARITY.
FT   DISULFID    308    323       BY SIMILARITY.
FT   DISULFID    326    330       BY SIMILARITY.
FT   DISULFID    503    512       BY SIMILARITY.
FT   DISULFID    507    520       BY SIMILARITY.
FT   DISULFID    523    532       BY SIMILARITY.
FT   DISULFID    536    552       BY SIMILARITY.
FT   DISULFID    555    569       BY SIMILARITY.
FT   DISULFID    559    577       BY SIMILARITY.
FT   DISULFID    580    589       BY SIMILARITY.
FT   DISULFID    593    614       BY SIMILARITY.
FT   DISULFID    617    625       BY SIMILARITY.
FT   DISULFID    621    633       BY SIMILARITY.
FT   MOD_RES    1162   1162       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1188   1188       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1258   1258       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1284   1284       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1062   1062       S -> N (IN REF. 3).
FT   CONFLICT   1080   1082       PYT -> SYR (IN REF. 3).
SQ   SEQUENCE   1308 AA;  146957 MW;  D944BB0996A08B41 CRC64;

  Query Match           41.6%;  Score 2833;  DB 1;  Length 1308;
  Best Local Similarity 43.9%;  Pred. No. 2e-144;
  Matches  594;  Conservative  194;  Mismatches  394;  Indels  170;  Gaps   31;

Qy          1 MELA-ALCRWGLLL--ALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQ 57
              |:||  |  || ||  |     :||  || ||: ||   :  |     ||  |: |:||
Db          1 MKLATGLWVWGSLLVAARTVQPSASQSVCAGTENKLSSLSDLEQQYRALRKYYENCEVVM 60

Qy         58 GNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVL 117
              ||||:|  :  |  ||||: |:|| ||||:| || | :||: |||:|||:|:|| ||||:
Db         61 GNLEITSIEHNRDLSFLRSIREVTGYVLVALNQFRYLPLENLRIIRGTKLYEDRYALAIF 120

Qy        118 DNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNN 177
                |      |          ||:|| |::|||| ||| : :|  ||| ||| |:||
Db        121 LNYRKDGNF---------GLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRNPW 171

Qy        178 QLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCC 236
              :||: |  |  |  |   | | |||| :   ||:|||||||  |  || ||  :|||
Db        172 PSNMTLVSTIGSSGCGRCHKSCTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCC 230

Qy        237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
              | :|| ||:||| :|| ||::|| || |   ||    ||  ||:   |  :||:|| ||
Db        231 HRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFCV 290

Qy        297 TACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF 356
              ||:|::  |  ||   ||    ||  |:| : |: |:  | : | |:|    : :
Db        291 KKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICPKACDGIGTGSLMSAQTV 348

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
                :|  | || |:| ||     ||| :    : ||:| || |: ||||:| |  ||
Db        349 DSSNIDKFINCTKINGNLIFLVTGIHGDPYNAIDAIDPEKLNVFRTVREITGFLNIQTWP 408

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
              :: | ||| ||  | ||:|::|   | |:  ||: |  :||:|: :|   |  |::||:
Db        409 PNMTDFSVFSNLVTIGGRVLYSGLSLLILKQQGITSLQFQSLKEISAGNIYITDNSNLCY 468

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
              ||: |  ||   :| ::   ||  : |  ||: |: ||:   |||||| ||::| :| |
Db        469 YHTINWTTLFSTVNQRIVIRDNRRAENCTAEGMVCNHLCSNDGCWGPGPDQCLSCRRFSR 528

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCFGPEADQCVACAHYKDPPF 595
              |: |:|  | :  |  ||: |   |: |  :|:  ::| :|| ||  | |  |:|:|| |
Db        529 GKICIESCNLYDGEFREFENGSICVECDSQCEKMEDGLLTCHGPGPDNCTKCSHFKDGPN 588

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----------PA 644
              || :||  ::   |:  |:|: |::  | ||  |||  |           |
Db        589 CVEKCPDVLQGANSF--IFKYADQDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQ 646

Qy        645 EQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPS 703
              | :||  | : |: |: ::|:: : | : ::|:  | :|  :|| | ||||||||||
Db        647 HAR-TPL--IAAGVIGGLFILVIMALTFAVYVRRKSIK-KKRALRRFL-ETELVEPLTPS 701

Qy        704 GAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWVPEGETVKIPVA-IKILNET 758
              | |||||:|||||||::||||||||||    ||   : ||      |:  : :::| |
Db        702 GTAPNQAQLRILKETELKRVKVLGSGAFG----TVYKGIWVPEGETVKIPVA-IKILNET 756

Qy        759 TSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLG 818
              | |||| | :||| :|| |  |:: ||||:||: |:|||||||:||||::| |::  :|
Db        757 TGPKANVEFMDEALIMASVDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIG 816

Qy        819 SQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYH 878
              || |||||:|||||| |||: ||||||||||||||||||||||||||||: || ||:
Db        817 SQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYN 876

Qy        879 ADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKG 938
              |||||:|||||||| |  |:|||||||||||||:|||||||  |||||||  ||||||||||
Db        877 ADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKG 936

Qy        939 ERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNED-LGPA 997
              |||||||||||||::|||||||:: ||:|:|| :|||||||||||::||| :| :
Db        937 ERLPQPPICTIDVYIVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQGDDRMKLP 996

Qy        998 SPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDP---------------------- 1035
              || || |:::||:::|: |::|||||||| | |  | |
Db        997 SPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTSRTRIDSNRSEIGHSPPPA 1055

Qy       1036 -APGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGA 1094
               | :|    ::    :|: |   : :    :    | :|:|  :|| :::||     |
Db       1056 YTPMSGSQFVYQDGGFATQQG---MPMPYTATTSTIPEAPVA--QGATAEMFDDSCCNGT 1110

Qy       1095 AKGLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRP 1147
               :       : |  |||| ||||  |         : :||: |:   |: ||:|  :
Db       1111 LRKPVVPHVQEDSSTQRYSADPTVFAPERNPRAELDEEGYMTPMHDKPKQEYLNPVE--- 1167

Qy       1148 QPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQP 1207
                 | |  : |                 ||| ::       |::|||| :   |
Db       1168 ------ENPFVSRR--------------KNGDLQ-------ALDNPEYHSASSG----- 1194

Qy       1208 HPPPA---------------------------------FSPAFDNLYYWDQDPPERGA- 1232
              ||  |                                 ||||  ||:   | |
Db       1195 -PPKAEDEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPDYWNHSLPPRSTL 1253

Qy       1233 -PPSTFKGTPT--------------AENPEYL 1249
               |   :  |               |||||||
Db       1254 QHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285
```

```
RESULT 8
XMRK_XIPMA
ID   XMRK_XIPMA     STANDARD;      PRT;  1167 AA.
AC   P13388;
DT   01-JAN-1990 (Rel. 13, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN   XMRK OR TU.
OS   Xiphophorus maculatus (Southern platyfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8083;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90015140; PubMed=2797166;
RA   Wittbrodt J., Adam D., Malitschek B., Maueler W., Raulf F.,
RA   Telling A., Robertson S.M., Schartl M.;
RT   "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT   inducing Tu locus in Xiphophorus.";
RL   Nature 341:415-421(1989).
RN   [2]
RP   REVISION TO 515.
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16891; CAA34770.2; -.
DR   PIR; S06142; S06142.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT   SIGNAL        1     25
FT   CHAIN        26   1167       MELANOMA RECEPTOR PROTEIN-TYROSINE
FT                                KINASE.
FT   DOMAIN        26    642       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     643    665       POTENTIAL.
FT   DOMAIN       666   1167       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       710    977       PROTEIN KINASE.
FT   NP_BIND      716    724       ATP (BY SIMILARITY).
FT   BINDING      743    743       ATP (BY SIMILARITY).
FT   ACT_SITE     835    835       BY SIMILARITY.
FT   DISULFID     195    204       BY SIMILARITY.
FT   DISULFID     199    212       BY SIMILARITY.
FT   DISULFID     220    228       BY SIMILARITY.
FT   DISULFID     224    236       BY SIMILARITY.
FT   DISULFID     237    245       BY SIMILARITY.
FT   DISULFID     241    253       BY SIMILARITY.
FT   DISULFID     256    265       BY SIMILARITY.
FT   DISULFID     269    296       BY SIMILARITY.
FT   DISULFID     300    311       BY SIMILARITY.
FT   DISULFID     315    330       BY SIMILARITY.
FT   DISULFID     333    337       BY SIMILARITY.
FT   DISULFID     504    513       BY SIMILARITY.
FT   DISULFID     508    521       BY SIMILARITY.
FT   DISULFID     524    533       BY SIMILARITY.
FT   DISULFID     537    553       BY SIMILARITY.
FT   DISULFID     556    569       BY SIMILARITY.
FT   DISULFID     560    577       BY SIMILARITY.
FT   DISULFID     593    615       BY SIMILARITY.
FT   DISULFID     618    626       BY SIMILARITY.
FT   DISULFID     622    634       BY SIMILARITY.
FT   CARBOHYD     114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1167 AA;  129934 MW;  4793E0749DC1D55A CRC64;

  Query Match             37.8%;  Score 2575.5;  DB 1;  Length 1167;
  Best Local Similarity   44.0%;  Pred. No. 1.1e-130;
  Matches  561;  Conservative  168;  Mismatches  392;  Indels  155;  Gaps   33;

Qy          4 AALCRWGLLLALLPPGAAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              ||| :  ||| |       ||    :|| ||  :: :    : |   :: :| || ||  |
Db          8 AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNVVLEN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
              ||:||   |  ||||| |||| |||||| |:|  :||  ||::||  |:| |: | |: |
Db         63 LEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLLVMSN 122

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                | ::|      ||::||| :||||| ||| :  || ||  :|| | ||  | :
Db        123 YQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDKTSNP 179

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
              : ||     | |  |   |    ||      ||  |: :||  |  ||:|| | |||:|
Db        180 TMNLIPHAFERQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              || |||||: :|||||  ||  | |:  ||     |:  : : : ||  :|||||:||
Db        240 HCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIG 358
              || ||: |: |:|   |      ||  |:| : |: |   | :|| |:|:  : :|:  :
Db        300 CPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSL-SNTIAVN 356

Qy        359 ITEL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPD 417
              | :  |: | || | :     ||:|||      : || |    |::|||||| |  ||:
Db        357 STNIRSFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWWPE 416

Qy        418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCF 476
              ::  |||||||::|||||   :  :| : :|   : |||||||:|: :|  :: :   | :
Db        417 NMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRY 476

Qy        477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
              :|: | :|||:  |:: : |          |   |:  |:  |||||| ||:|    |
Db        477 ANTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCVSCLHVDR 529

Qy        537 GQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFC 596
```

```
                 |  ||  | :||| |||      |: || ||  |  |:||:||    |   ||::| | |
Db        530 GGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKSAHFQDGPQC 589

Qy        597 VARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVS 656
              : ||| |:  |   : |||: |: | |||| ||| |      || :  |  :  |
Db        590 IPRCPHGILGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGD-IVSHSSLAVG 647

Qy        657 AVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
              | |:|: |:: ::  :|::||: | || |:| |||| |||||||||| |||| :|||
Db        648 LVSGLLITVIVALLIVVLLRRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQAFLRILK 706

Qy        717 ETELRKVKVLGSGAFGFNNFTV--SFW------LRVPKVSASHLEVLRENTSPKANKEIL 768
              ||| :| :|||||||    ||    |      :|:|      ::|||| |||| |:|:|
Db        707 ETEFKKDRVLGSGAFG----TVYKGLWNPDGENIRIPVA----IKVLREATSPKVNQEVL 758

Qy        769 DEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQ 828
              ||||||| |  |:| |||||||| |||||||||||||:||::: |:  | ||||:|
Db        759 DEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQWLLNWCVQ 818

Qy        829 IAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKW 888
              |||||:|||:  |||||||||||:|:||||||||||::||  || || ||||||||||
Db        819 IAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPIKW 878

Qy        889 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICT 948
              |||||||:  :|||||||||||||||||:||||||||:||  :|| ||||||||||||
Db        879 MALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERLPQPPICT 938

Qy        949 IDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1008
              |:||||::||||||   |||||||| |||:|||| |::|||       |  |   :  |
Db        939 IEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSLSDRRLFSRL 995

Qy       1009 LEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEE 1068
              |  ||  |:|||:|||:| :                    | |
Db        996 LSSDD--DVVDADEYLLPYKRI--------------------NRQGS------------ 1020

Qy       1069 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDGY 1127
                 |  |  |                    |  :  |   |   |||   |  : ||:
Db       1021 ----EPCIPPTGH-------------------PVRENSITLRNISDPTQNALEKDLDGH 1056

Qy       1128 VAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATLERAKTLS 1174
                 ||||||            |:  :     ||:  ::: |      | : |:
Db       1057 ----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNFSRPEYLN 1106

Qy       1175 PGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPP 1234
                :| :        |::::|:|   | |           :  ||          |: ||
Db       1107 TNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------LPQTGALT 1141

Qy       1235 STFKGTPTAENPEYLG 1250
                 | ||| ||||
Db       1142 GNGMFLPAAENLEYLG 1157
```

```
RESULT 9
ERB3_HUMAN
ID   ERB3_HUMAN     STANDARD;      PRT;  1342 AA.
AC   P21860;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN   ERBB3 OR HER3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90083234; PubMed=2687875;
RA   Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT   "Isolation and characterization of ERBB3, a third member of the
RT   ERBB/epidermal growth factor receptor family: evidence for
RT   overexpression in a subset of human mammary tumors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90311312; PubMed=2164210;
RA   Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA   Todaro G.J., Shoyab M.;
RT   "Molecular cloning and expression of an additional epidermal growth
RT   factor receptor-related gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN   [3]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RC   TISSUE=Placenta;
RX   MEDLINE=93282822; PubMed=7685162;
RA   Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT   "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT   tyrosine kinase.";
RL   Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC       SECRETED (SHORT FORM).
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC       SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC   -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M29366; AAA35790.1; -.
DR   EMBL; M34309; AAA35979.1; -.
DR   EMBL; S61953; AAB26935.1; -.
DR   PIR; A36223; A36223.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3431; ERBB3.
DR   MIM; 190151; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1342       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN      20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   644    664       POTENTIAL.
FT   DOMAIN     665   1342       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     709    966       PROTEIN KINASE.
FT   NP_BIND    715    723       ATP (BY SIMILARITY).
```

```
FT   BINDING     742    742       ATP (BY SIMILARITY).
FT   ACT_SITE    834    834       BY SIMILARITY.
FT   DISULFID    186    194       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    552    565       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    141    183       EILSGGVYIEKNDKLCHMDTIDWRDIVRDRDAEIVVKDNGR
FT                                SC -> GQFPMVPSGLTPQPAQDWYLLDDDPRLLTLSASSK
FT                                VPVTLAAV (IN SHORT ISOFORM).
FT   VARSPLIC    184   1342       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    560    560       E -> G (IN REF. 2).
FT   CONFLICT   1064   1064       E -> G (IN REF. 2).
SQ   SEQUENCE   1342 AA;  148097 MW;  7201E7F66CA374BD CRC64;

  Query Match             34.0%;  Score 2316.5;  DB 1;  Length 1342;
  Best Local Similarity   39.5%;  Pred. No. 9.5e-117;
  Matches  520;  Conservative  196;  Mismatches  459;  Indels  143;  Gaps   36;

Qy         10 GLLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT 67
              ||| :|       |  || ||   |  :    |     |  ||: |:|| ||||:
Db         11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

Qy         68 NASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
              || ||||| |:|| ||||:| |:   :||  ||:||||:::  :|: |:     ||  |
Db         71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVM-----LNYNT 125

Qy        128 PVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTN 187
                  |   ||:|:|  ||||| ||| |::| :||: ||| |:||    :     ::  :
Db        126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy        188 RSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCAAGCTG 246
                |:| ||  :||| |||  ||||:||:|:||  |   | || |  |||::|| ||:|
Db        179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNGHCFGPNPNQCCHDECAGGCSG 237

Qy        247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
              |: :|| || ||| || |   ||  : ||  ||:  ||| :| :|  || :||:|::
Db        238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFQLEPNPHTKYQYGGVCVASCPHNFV-V 296

Qy        307 DVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKF--IGITELE- 363
              |  ||   ||    ||  ::| : || |   | : | | |     : |:|  :  : ::
Db        297 DQTSCVRACPPDKMEVD-KNGLKMCEPCGGLCPKACEGTG-----SGSRFQTVDSSNIDG 350

Qy        364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423
              |  | || |:| ||    :|||       | ||:| || |: |||||| | :||  : : |
Db        351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVFRTVREITGYLNIQSWPPHMHNFS 410

Qy        424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
              || ||  | || |:|  :| | :: | :: || |||:|: :|   |  |  ||: |:: |
Db        411 VFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNW 470

Qy        483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECV 541
               :: | | :  |     |||  :|| ||  |  ||: | |||||| ||::|  : ||  ||
Db        471 TKVLRGPTEERLDIKHNRPRRDCVAEGKVCDPLCSSGGCWGPGPGQCLSCRNYSRGGVCV 530

Qy        542 EECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCP 601
                |  | |||: :   |  |||||||  |: || |  :| |  ||::| | ||: ||
Db        531 THCNFLNGEPREFAHEAECFSCHPECQPMEGTATCNGSGSDTCAQCAHFRDGPHCVSSCP 590

Qy        602 SGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRA----SPLTSIVSA 657
               ||    :  ||:|:|| :  |:||  ||| |   : : |  :       : ||  ::
Db        591 HGVLG--AKGPIYKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLIGKTHLTMALTV 648

Qy        658 VVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
              : |  |||:  :: |  :  | ::|: |  ||| |:  | :|||  ||    |:   || |
Db        649 IAG--LVVIFMMLGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705

Qy        717 ETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASH----LEVLRENTSPKANKEILDEAY 772
              ||||||:|||||| ||    ||   : :|:  :      ::|: : :  :: : : |
Db        706 ETELRKLKVLGSGVFG----TVHKGVWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHML 761

Qy        773 VMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKG 832
               :  :   :: ||||:|  |::|||||  :| | ||||||::|| || | |||| :|||||
Db        762 AIGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVRQHRGALGPQLLLNWGVQIAKG 821

Qy        833 MSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALE 892
              | |||:  :|||:|||||||:|||: |:: |||:| || |: :      | ||||||||
Db        822 MYYLEEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKWMALE 881

Qy        893 SILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVY 952
              ||   ::|||||||||||||||||||:|| |:   |:|||||||||| || |||||||
Db        882 SIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEVPDLLEKGERLAQPQICTIDVY 941

Qy        953 MIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPA---SPLDSTFYRSLL 1009
              |:|||||||    || |:|| :||:|||||| |::||: |  ||      |          |
Db        942 MVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKL 1000

Qy       1010 EDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP-SEE 1068
              |: ::   :| :  |  ::                    : :| : |  |:| :   :
Db       1001 EEVELEPELDLDLDLEAEED------------------NLATTTLGSALSLPVGTLNRP 1041

Qy       1069 EAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHD-PSPLQRYSEDPTVPLP------ 1121
                :| |:|| |   :  |:||   :  |  :  | |:  :      |:|
Db       1042 RGSQSLLSPSSGY-MPMNQGNLGESCQESAVSGSSERCPRPVSLH------PMPRGCLAS 1094

Qy       1122 SETDGYVA----------------PLTCSPQPE----YVNQPDVRPQPPSPREGP----- 1156
                ::|:|                : ||:|     | :|      | :|   |
Db       1095 ESSEGHVTGSEAELQEKVSMCRSRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPPGLEEE 1154

Qy       1157 ------LPAARPAGATLERAKTLSP-GKNGVV------KDVFAFGGAVENPEYLTPQGGA 1203
                   :|     |    | |||  | : |:      :|      |  ||:  :
Db       1155 DVNGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEEEDED--------EEYEYMNRRRRH 1206

Qy       1204 APQPHPPPAFSPAFDNLYYWD------------QDPPERGAPPSTFKGTPTAENPEYL 1249
                :| ||||   |      | |           |  |   |    ||   |: ||:
Db       1207 SP-PHPPRPSSLEELGYEYMDVGSDLSASLGSTQSCPLHPVPIMPTAGTTPDEDYEYM 1263


RESULT 10
ERB3_RAT
ID   ERB3_RAT       STANDARD;      PRT;  1339 AA.
AC   Q62799; Q62955;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE   (c-erbB3).
GN   ERBB3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96096535; PubMed=8522190;
RA   Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT   "Cloning of the rat ErbB3 cDNA and characterization of the
RT   recombinant protein.";
RL   Gene 165:279-284(1995).
RN   [2]
RP   REVISIONS TO 85; 513 AND 565.
RA   Hellyer N.J., Koland J.G.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 922-1097 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX   MEDLINE=97184212; PubMed=9030624;
RA   Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT   "Expression of neuregulins and their putative receptors, ErbB2 and
RT   ErbB3, is induced during Wallerian degeneration.";
RL   J. Neurosci. 17:1642-1659(1997).
CC   -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC       SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC       AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29339; AAC28498.2; -.
DR   EMBL; U52530; AAC53050.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1339       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT   DOMAIN       20    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    644    662       POTENTIAL.
FT   DOMAIN      663   1339       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      183    259       CYS-RICH.
FT   DOMAIN      707    964       PROTEIN KINASE.
FT   NP_BIND     713    721       ATP (BY SIMILARITY).
FT   BINDING     740    740       ATP (BY SIMILARITY).
FT   ACT_SITE    832    832       BY SIMILARITY.
FT   DISULFID    186    194       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    210    218       BY SIMILARITY.
FT   DISULFID    214    226       BY SIMILARITY.
FT   DISULFID    227    235       BY SIMILARITY.
FT   DISULFID    231    243       BY SIMILARITY.
FT   DISULFID    246    255       BY SIMILARITY.
FT   DISULFID    259    286       BY SIMILARITY.
FT   DISULFID    290    301       BY SIMILARITY.
FT   DISULFID    305    320       BY SIMILARITY.
FT   DISULFID    323    327       BY SIMILARITY.
FT   DISULFID    500    509       BY SIMILARITY.
FT   DISULFID    504    517       BY SIMILARITY.
FT   DISULFID    520    529       BY SIMILARITY.
FT   DISULFID    533    549       BY SIMILARITY.
FT   DISULFID    556    573       BY SIMILARITY.
FT   DISULFID    576    585       BY SIMILARITY.
FT   DISULFID    589    610       BY SIMILARITY.
FT   DISULFID    613    621       BY SIMILARITY.
FT   DISULFID    617    629       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    408    408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    522    522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1028   1028       L -> P (IN REF. 3).
SQ   SEQUENCE   1339 AA;  147545 MW;  0AA5F2402BBFDF1E CRC64;

  Query Match           32.9%;  Score 2244.5;  DB 1;  Length 1339;
  Best Local Similarity 39.7%;  Pred. No. 6.8e-113;
  Matches  512;  Conservative  173;  Mismatches  437;  Indels  167;  Gaps   37;

Qy          3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGN 59
              |  ||     |  |  |:    |  || ||   | :   :      |  ||: |:|| ||
Db          7 LQVLC----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLYKLYEKCEVVMGN 62

Qy         60 LELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
               ||:     || ||||| |:|| ||||:| |:   :||  ||:|||||::   :|: |:
Db         63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLPNLRVVRGTQVYDGKFAIFVM-- 120

Qy        120 GDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQL 179
                 ||  |      |   ||:|:   ||||| ||| |:| :||: ||| |:||
Db        121 ---LNYNT----NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDWRDIVRVR--- 170

Qy        180 ALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHE 238
              ::  |    | ||  :||| |||   :||| ||:|:||  |  || || |  |||:
Db        171 GAEIVVKNNGANCPPCHEVCKG-RCWGPGPDDCQILTKTICAPQCNGRCFGPNPNQCCHD 229

Qy        239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
              :|| ||:||: :|| ||  || || |   ||  : ||  ||:  |||  :| :|  || :
Db        230 ECAGGCSGPQDTDCFACRRFNDSGACVPRCPEPLVYNKLTFQLEPNPHTKYQYGGVCVAS 289

Qy        299 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGL--GMQYIKANSKF 356
              ||:|::  |    |   ||   ||  : | : || |   | : | |   |  :|   :|
Db        290 CPHNFV-VDQTFCVRACPPDKMEVD-KHGLKMCEPCGGLCPKACEGTGSGSRYQTVDSSN 347

Qy        357 IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWP 416
              |      | | || |:| ||    : ||      | ||:| || |: |||||| | :||
Db        348 ID----GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLNVFRTVREITGYLNIQSWP 403

Qy        417 DSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIHHNTHLC 475
              : : ||| ||  | || |:|  :| | :: | :: || |||:|: :|   |  | ||
Db        404 PHMHNFSVFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLC 463

Qy        476 FVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQF 534
              : |:: | :| | | :  |    :|| ||: ||  |  ||: | |||||| ||::|  :
Db        464 YHHSLNWTRLLRGPSEERLDIKYDRPLGECLAEGKVCDPLCSSGGCWGPGPGQCLSCRNY 523

Qy        535 LRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPP 594
```

```
                 |   ||  |  ||| |||:|:    |  ||||| |  |: || |  :| |  |||::| |
Db        524 SREGVCVTHCNFLQGEPREFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAHFRDGP 583

Qy        595 FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSC--VDLDDKGCPAEQRASPLT 652
              ||  || |:    :  ||:|:|| :  |:||  |||  |   :| |    ||   |
Db        584 HCVNSCPHGILG--AKGPIYKYPDAQNECRPCHENCTQGCNGPELQDCLGQAEVLMSKPH 641

Qy        653 SIVSAVVGILLVVVLGVVFGILIKRRQQKIR-KYTMRRLLQETELVEPLTPSGAMPNQAQ 711
               :::  ||  | |:| :: |  :  | ::|: |  ||| |:  | :||| ||    |:
Db        642 LVIAVTVG--LAVILMILGGSFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVL 698

Qy        712 MRILKETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASH----LEVLRENTSPKANKEI 767
              || |||||||:|||||| ||    ||   : :|:  :      ::|: : :  :: : :
Db        699 ARIFKETELRKLKVLGSGVFG----TVHKGIWIPEGESIKIPVCIKVIEDKSGRQSFQAV 754

Qy        768 LDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCM 827
              |    :  :   :: ||||:|  |::|||| :| | |||||:::|  || | |||| :
Db        755 TDHMLAVGSLDHAHIVRLLGLCPGSSLQLVTQYLPLGSLLDHVKQHRETLGPQLLLNWGV 814

Qy        828 QIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIK 887
              |||||| |||:  :|||||| |||::|||: |:: |||:| ||  |: :       | |||
Db        815 QIAKGMYYLEEHSMVHRDLALRNVMLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIK 874

Qy        888 WMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPIC 947
              |||||||   ::|||||||||||||||||||:|| |:   ||||||||||| || ||
Db        875 WMALESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPQIC 934

Qy        948 TIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRS 1007
              ||||||:|||||||   || |:|| :||:|||||| |::||:    || :|       |
Db        935 TIDVYMVMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAEPS 991

Qy       1008 LLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSE 1067
              :|   ::      |  | |:                              || | ||  |
Db        992 VLTTKEL-----QEAELEPEL----------------------------DLDLDLEAEE 1017

Qy       1068 E--------------------EAPRSPLAPSEG--------AGSDVFDGDLGMGAAKGLQ 1099
              |                    :| |:|| |         |    |  :  |  :  :
Db       1018 EGLATSLGSALSLPTGTLTRPRGSQSLLSPSSGYMPMNQSSLGEACLDSAVLGGREQFSR 1077

Qy       1100 SLPTHDPSPLQRYSEDPTVPLPSETDGYV----APL-----TC-------SPQPE----Y 1139
               :  | | |  |       |    ::|:|    | |       |       ||:|     |
Db       1078 PISLH-PIPRGR-------PASESSEGHVTGSEAELQEKVSVCRSRSRSRSPRPRGDSAY 1129

Qy       1140 VNQPDVRPQPPSPREGP-----------LPAARPAGATLERAKTLSP-GKNGVV------ 1181
              :|      | :|   |           :|     ||:  |  |||  | : |:
Db       1130 HSQRHSLLTPVTPLSPPGLEEEDGNGYVMPDTHLRGASSSREGTLSSVGLSSVLGTEEED 1189

Qy       1182 KDVFAFGGAVENPEYLTPQGGAAPQPHPP 1210
              :|       |  ||:  :   :| | ||
Db       1190 ED--------EEYEYMNRKRRGSP-PRPP 1209


RESULT 11
EGFR_DROME
ID   EGFR_DROME     STANDARD;      PRT;  1426 AA.
AC   P04412; O61601; Q9W2G0; P81868;
DT   13-AUG-1987 (Rel. 05, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE   (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN   EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX   MEDLINE=94350209; PubMed=8070664;
RA   Clifford R., Schupbach T.;
RT   "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT   that several genetically defined classes of alleles cluster in
RT   subdomains of the receptor protein.";
RL   Genetics 137:531-550(1994).
RN   [2]
RP   REVISIONS.
RA   Clifford R., Schupbach T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=85124611; PubMed=2982499;
RA   Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT   "The Drosophila EGF receptor gene homolog: conservation of both
RT   hormone binding and kinase domains.";
RL   Cell 40:599-607(1985).
RN   [4]
RP   SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC   STRAIN=Oregon-R; TISSUE=Embryo;
RX   MEDLINE=87002474; PubMed=3093080;
RA   Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT   "Alternative 5' exons and tissue-specific expression of the
RT   Drosophila EGF receptor homolog transcripts.";
RL   Cell 46:1091-1101(1986).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=99102120; PubMed=9882502;
RA   Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT   "Several levels of EGF receptor signaling during photoreceptor
RT   specification in wild-type, Ellipse, and null mutant Drosophila.";
RL   Dev. Biol. 205:129-144(1999).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [7]
RP   SEQUENCE OF 959-1078 FROM N.A.
```

```
RC   STRAIN=Daekwanryeong;
RX   MEDLINE=85137938; PubMed=2983232;
RA   Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT   "A Drosophila genomic sequence with homology to human epidermal
RT   growth factor receptor.";
RL   Nature 314:178-180(1985).
RN   [8]
RP   SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP   ANALYSIS.
RX   MEDLINE=92038942; PubMed=1936959;
RA   Raz E., Schejter E.D., Shilo B.Z.;
RT   "Interallelic complementation among DER/flb alleles: implications for
RT   the mechanism of signal transduction by receptor-tyrosine kinases.";
RL   Genetics 129:191-201(1991).
RN   [9]
RP   REVIEW.
RX   MEDLINE=97248481; PubMed=9094709;
RA   Perrimon N., Perkins L.A.;
RT   "There must be 50 ways to rule the signal: the case of the Drosophila
RT   EGF receptor.";
RL   Cell 89:13-16(1997).
CC   -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC       WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC       MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC       CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC       DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC       POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC       ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC       AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC       SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC       CUTICLE.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC       MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC       PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC       TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC       UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC       TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC       EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC       IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC       POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC       AND THORACIC AND ABDOMINAL GANGLIA.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF052754; AAC08536.1; -.
DR   EMBL; AF052753; AAC08536.1; JOINED.
DR   EMBL; AF052754; AAC08535.1; -.
DR   EMBL; AF052752; AAC08535.1; JOINED.
DR   EMBL; K03054; AAA51462.1; -.
DR   EMBL; K03417; AAA51460.1; -.
DR   EMBL; K03416; AAA50965.1; -.
DR   EMBL; K03418; AAA51461.1; -.
DR   EMBL; AF109077; AAD26134.1; -.
DR   EMBL; AF109078; AAD26132.1; -.
DR   EMBL; AF109082; AAD26132.1; JOINED.
DR   EMBL; AF109078; AAD26133.1; -.
DR   EMBL; AF109084; AAD26133.1; JOINED.
DR   EMBL; AF109079; AAD26130.1; -.
DR   EMBL; AF109081; AAD26130.1; JOINED.
DR   EMBL; AF109079; AAD26131.1; -.
DR   EMBL; AF109083; AAD26131.1; JOINED.
DR   EMBL; AF109080; AAD26135.1; -.
DR   EMBL; AE003454; AAF46732.1; -.
DR   EMBL; X02293; CAA26157.1; -.
DR   EMBL; X78920; CAA55523.1; -.
DR   EMBL; X78918; CAA55521.1; -.
DR   EMBL; X78919; CAA55522.1; -.
DR   PIR; A00640; GQFFE.
DR   HSSP; P11362; 1FGK.
DR   FlyBase; FBgn0003731; Egfr.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 7.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW   Developmental protein.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1426       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    868       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    869    889       POTENTIAL.
FT   DOMAIN      890   1426       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      938   1198       PROTEIN KINASE.
FT   NP_BIND     944    952       ATP (BY SIMILARITY).
FT   BINDING     971    971       ATP (BY SIMILARITY).
FT   ACT_SITE   1063   1063       BY SIMILARITY.
FT   MOD_RES     902    902       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

  Query Match           26.9%;  Score 1836;  DB 1;  Length 1426;
  Best Local Similarity   31.8%;  Pred. No. 5.5e-91;
  Matches  456;  Conservative  183;  Mismatches  436;  Indels  360;  Gaps   46;

Qy         24 QVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPT-NASLSFLQDIQEVQG 82
              ::| ||  :| :|:: | |   ||  |  | | |||:||:||  |  |||| :|:|| |
Db        100 KICIGTKSRLSVPSNKEHHYRNLRDRYTNCTYVDGNLKLTWLPNENLDLSFLDNIREVTG 159

Qy         83 YVLIAHNQVRQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGL 137
              |:||:|  |::|   :|:|:||  ||     |: ||| |                :  :
Db        160 YILISHVDVKKVVFPKLQIIRGRTLFSLSVEEEKYALFV---------------TYSKM 203

Qy        138 RELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSP 197
                |::  | ::| | |    | ||:  || | :|              |   | | |
Db        204 YTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFTAPERECPKCHE 263

Qy        198 MCKGSRCWGESSEDCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDCLAC 255
              |   ||||  ::||  ::  |:  ||  || || | :||| || ||||   ||:||
Db        264 SCTHG-CWGEGPKNCQKFSKLTCSPQCAGGRCYGPKPRECCHLFCAGGCTGPTQKDCIAC 322

Qy        256 LHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVC 315
              :|    : :  || :  ||  |:    ||||:| :||:||  ||  :|  | |:|   |
Db        323 KNFFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACVRSC 381

Qy        316 PLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITEL------EFAGCKK 369
              |    :  |     | |: || : |               |:| |       | |
Db        382 PQDKMDKGGE-----CVPCNGPCPKTC-------------PGVTVLHAGNIDSFRNCTV 422

Qy        370 IFGSLAFLPESFDG--DPASNTA------PLQPEQLQVFETLEEITGYLYISAWPDSLPD 421
              | |::  | ::| |  |  :|         || ||: :|| |::|||||| |       :
Db        423 IDGNIRILDQTFSGFQDVYANYTMGPRYIPLDPERREVFSTVKEITGYLNIEGTHPQFRN 482

Qy        422 LSVFQNLQVIRGRILHNGAY-SLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              || |:||: | || |    : :| :    :  | :|:|::: ||  :| ||  ||:|  :
Db        483 LSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRDLCYVSNI 542

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
```

```
                 |  : : | | :    |   | |   |  |   |    |||| |  ||: |  |     |
Db        543 RWPAIQKEPEQKVWVNENLRADLCEKNGTICSDQCNEDGCWGAGTDQCLTCKNFNFNGTC 602

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              : :|  :    :   : | |  |||||:      || |  || |  | | :|   ||: |
Db        603 IADCGYISNAYK--FDNRTCKICHPECR------TCNGAGADHCQECVHVRDGQHCVSEC 654

Qy        601 P--------------------SGVK--------------------------------PDL 608
              |                    :| |                                ||
Db        655 PKNKYNDRGVCRECHATCDGCTGPKDTIGIGACTTCNLAIINNDATVKRCLLKDDKCPD- 713

Qy        609 SYMPIWKF--PDEEGACQP-------------CPI------------NCTH--------- 632
              |   |::  | |:|: :|             | :            |||
Db        714 GY--FWEYVHPQEQGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCET 771

Qy        633 ---------------------------------SC-----VDLDDKG------------ 641
                                                ||     | ::  |
Db        772 ECPADHYTDEEQRECFQRHPECNGCTGPGADDCKSCRNFKLFDANETGPYVNSTMFNCTS 831

Qy        642 -CPAEQR--------------ASPLTS-------------IVSAVVGILLVVVLGVVFGI 673
               || | |              |||  |             |::  | :  : :| ||
Db        832 KCPLEMRHVNYQYTAIGPYCAASPPRSSKITANLDVNMIFIITGAVLVPTICILCVV--T 889

Qy        674 LIKRRQQKIRKYT--MRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAF 731
              | |::|| :| |  |    |  |  ||| ||     |  ::||:|: ||||  ||| |||
Db        890 YICRQKQKAKKETVKMTMALSGCEDSEPLRPSNIGANLCKLRIVKDAELRKGGVLGMGAF 949

Qy        732 GFNNPTVSFWLRVP-----KVSASHLEVLRENTSPKANKEILDEAYVMAGVGSPYVSRLL 786
              |       |  ||     |:  :  |:|: :|  ::::| | |||:||      : :||
Db        950 G--RVYKGVW--VPEGENVKIPVAIKELLK-STGAESSEEFLREAYIMASEEHVNLLKLL 1004

Qy        787 GICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDL 846
              :|::| : |:|||| ||||:|| || ::||: |||| ||||||||||: ||||||
Db       1005 AVCMSSQMMLITQLMPLGCLLDYVRNNRDKIGSKALLNWSTQIAKGMSYLEEKRLVHRDL 1064

Qy        847 AARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVW 906
              |||||||::|: ||||||||:||  |  || | |||:||||:||| |  | || :||||
Db       1065 AARNVLVQTPSLVKITDFGLAKLLSSDSNEYKAAGGKMPIKWLALECIRNRVFTSKSDVW 1124

Qy        907 SYGVTVWELMTFGAKPYDGIPARBIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECR 966
              ::|||:|||:||| :|:: |||::||||:| | :| || ||::|:|  :: || :|:  |
Db       1125 AFGVTIWELLTFGQRPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCWHLDAAMR 1184

Qy        967 PRFRELVSEFSRMARDPQRFVVIQNEDLG--PASPLDSTFYRSLLEDD---DMGDLVDAE 1021
              | |::| : |:  |||| |::  |  :     ||        | |  | |    :   |
Db       1185 PTFKQLTTVFAEFARDPGRYLAIPGDKFTRLPA-------YTSQDEKDLIRKLAPTTDGS 1237

Qy       1022 EYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAP------RSPL 1075
              | :    :  |  |||         ||:  |              :| |      : |
Db       1238 EAIAKPDDYLQPKAAPGPS------HRTDCT--------------DEMPKLNRYCKDPS 1276

Qy       1076 APSEGAGSDVFDG---DLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLT 1132
                :   | |  |    ::|:|  :                     : || : | |: | |
Db       1277 NKNSSTGDDERDSSAREVGVGNLR--------------------LDLPVDEDDYLMP-T 1314

Qy       1133 CSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVE 1192
              | | |    |  :      :| :  : |   |   :              |:     :|:
Db       1315 CQPGPNNNNNMN------NPNQNNMAAVGVAAGYM--------------DLIGVPVSVD 1353

Qy       1193 NPEYL----TPQGGAAPQPH-----------------PPPAFSP-AFDNLYYWD 1224
              |||||    |   | :| |                 | |   | : |: || |
Db       1354 NPEYLLNAQTLGVGESPIPTQTIGIPVMGGPGTMEVKVPMPGSEPTSSDHEYYND 1408


RESULT 12
ERBB_ALV
ID   ERBB_ALV       STANDARD;      PRT;   634 AA.
AC   P00534;
DT   21-JUL-1986 (Rel. 01, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian leukosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11864;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85228222; PubMed=2988784;
RA   Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA   Crittenden L.B., Raines M.A., Kung H.-J.;
RT   "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT   processing and promoter insertion result in expression of an
RT   amino-truncated EGF receptor.";
RL   Cell 41:719-726(1985).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC       IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC       AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC       THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC       C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC       PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10066; AAA48763.1; ALT_INIT.
DR   PIR; A00643; TVCHLV.
DR   PIR; B00643; TVFVLV.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
SQ   SEQUENCE   634 AA;  70891 MW;  E705E33A0BE01FCC CRC64;

  Query Match             24.0%;  Score 1638.5;  DB 1;  Length 634;
  Best Local Similarity   50.2%;  Pred. No. 8.3e-81;
  Matches  357;  Conservative   80;  Mismatches  143;  Indels  131;  Gaps   19;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
              |||: | | ||  ||:||  :   : :||: |    || |  |||  |     :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705
                |   || : || |:| :||:|:  |: ::||   :|| |:||||| ||||||||||
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSP 761
              |||| :|||||| :|||||||||    ||   | :|:     : |    ||| |||
Db        118 APNQAHLRILKETEFKKVKVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSP 173

Qy        762 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 821
              |||||||||||||| | :|:| ||||||||||||:||||||||||::||::  :|||
Db        174 KANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQY 233

Qy        822 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 881
```

```
                |||||:||||||:|||: ||||||||||||||||:| ||||||||||:||  || ||||:|
Db        234 LLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEG 293

Qy        882 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 941
              |||||||||||||| | :|||||||||||||||||:|||||||| ||  :|||||||
Db        294 GKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERL 353

Qy        942 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPL 1000
              |||||||||||||||||||:: ||:||||::|||:||||| |::||| :| :   ||
Db        354 PQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPT 413

Qy       1001 DSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLT 1060
              || |||:|:|::|| |:|||:||||| ||||                | ||
Db        414 DSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-------- 449

Qy       1061 LGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSED 1115
                        |:||     | |  : ::  |        :  |  |  : | :|||| |
Db        450 ----------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSD 491

Qy       1116 PTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTL 1173
              ||      |:  ||::      | ||||||  : |: ||
Db        492 PTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPS--------------------T 522

Qy       1174 SPGKNGVVKDVF---------------AFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFD 1218
              :  :| :  ::                :   ||:|||||          :  |      |:
Db        523 AMVQNQIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAKTVFE 574

Qy       1219 NLYYWDQ--------DPPE-------RGAPPSTFKGTPTAENPEYLGLDVP 1254
              :  || |        | |:          |:     | ||||||| :  |
Db        575 SSPYWIQSGNHQINLDNPDYQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 625


RESULT 13
EGFR_CHICK
ID   EGFR_CHICK     STANDARD;      PRT;   703 AA.
AC   P13387;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE   (Fragment).
GN   EGFR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88261272; PubMed=3260329;
RA   Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA   Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT   "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT   expression in mouse cells, and differential binding of EGF and
RT   transforming growth factor alpha.";
RL   Mol. Cell. Biol. 8:1970-1978(1988).
CC   -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC       AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC       VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC       dimerization, internalization of the EGF-receptor complex,
CC       induction of the tyrosine kinase activity, stimulation of cell DNA
CC       synthesis, and cell proliferation.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20386; AAA48760.1; -.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 4.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
KW   Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW   Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   SIGNAL        1     30
FT   CHAIN        31   >703       EPIDERMAL GROWTH FACTOR RECEPTOR.
FT   DOMAIN       31    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    667       POTENTIAL.
FT   DOMAIN      668   >703       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    197    206       BY SIMILARITY.
FT   DISULFID    201    214       BY SIMILARITY.
FT   DISULFID    222    230       BY SIMILARITY.
FT   DISULFID    226    238       BY SIMILARITY.
FT   DISULFID    239    247       BY SIMILARITY.
FT   DISULFID    243    255       BY SIMILARITY.
FT   DISULFID    258    267       BY SIMILARITY.
FT   DISULFID    271    298       BY SIMILARITY.
FT   DISULFID    302    314       BY SIMILARITY.
FT   DISULFID    318    333       BY SIMILARITY.
FT   DISULFID    336    340       BY SIMILARITY.
FT   DISULFID    513    522       BY SIMILARITY.
FT   DISULFID    517    530       BY SIMILARITY.
FT   DISULFID    533    542       BY SIMILARITY.
FT   DISULFID    546    562       BY SIMILARITY.
FT   DISULFID    565    581       BY SIMILARITY.
FT   DISULFID    569    589       BY SIMILARITY.
FT   DISULFID    592    601       BY SIMILARITY.
FT   DISULFID    605    627       BY SIMILARITY.
FT   DISULFID    630    638       BY SIMILARITY.
FT   DISULFID    634    646       BY SIMILARITY.
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    200    200       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    359    359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    420    420       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    633    633       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    648    648       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     703    703
SQ   SEQUENCE   703 AA;  77427 MW;  AFF2DE11B735A690 CRC64;

  Query Match             23.4%;  Score 1595;  DB 1;  Length 703;
  Best Local Similarity   44.6%;  Pred. No. 2e-78;
  Matches  316;  Conservative  111;  Mismatches  251;  Indels   30;  Gaps   14;

Qy          8 RWGLLLALLPPGAA------STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE 61
                |   :| ||  | |         :|| ||: ||       | |    |: :|  |:||  |||
Db         13 RGAAVLVLLLLGVALCSAVEEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNNCEVVLSNLE 72

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
               :||:  |  |:||: |||| |||||| | |  :||: |:|:||  |:::::||||| |
Db         73 ITYVEHNRDLTFLKTIQEVAGYVLIALNMVDVIPLENLQIIRGNVLYDNSFALAVLSNYH 132

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
               :| |          ||||| :: |:||| ||| |  ||:||  ||:|| ||   : :  |
Db        133 -MNKTQ--------GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy        182 TLID-TNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
```

```
          |::|  :   :|  | | |    |||   ::||:||: :||  |: ||:| :|:|||| |
Db        183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQQCSGRCRGKVPSDCCHNQ 242

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              ||||||||: ||||||  |     |:  || || ||  |::   ||||:|:|||:||  |
Db        243 CAAGCTGPRESDCLACRKFRDDATCKDTCPPLVLYNPTTYQMDVNPEGKYSFGATCVREC 302

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              |:||: || |||   |     ||  |:| ::|:||   |::|| |:|:  :|     |
Db        303 PHNYVVTDHGSCVRSCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  | || | :: || :| ||  : | || |::| || |::||:|:| | ||||:
Db        361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
                ||  |:||::|||   :| ||| :  | |  ||||||:|:  |   |  | :||:
Db        421 ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 480

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: |  ||    |    || :::|  :   |  ||:  ||||||  | :|  | | :
Db        481 TMNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPGPFHCFSCRFFSRQK 540

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNG---SVTCFGPEADQCVACAHYKDPPF 595
              |||::| :||| |||:     ||||| ||  ||    : || ||  | |: |||: | |
Db        541 ECVKQCNILQGEPREFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCMKCAHFIDGPH 600

Qy        596 CVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIV 655
              ||  ||:||  :   : :||: |    || |  ||| |     :|||     |   ||
Db        601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656

Qy        656 SAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTP 702
              : || |:| :||:|:  |: ::||   :|| |:|||||| ||||||||
Db        657 AGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTP 703
```

```
RESULT 14
ERBB_AVIER
ID   ERBB_AVIER     STANDARD;      PRT;   604 AA.
AC   P00535;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ES4).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=79685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H;
RX   MEDLINE=84026539; PubMed=6313229;
RA   Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT   "The erbB gene of avian erythroblastosis virus is a member of the src
RT   gene family.";
RL   Cell 35:71-78(1983).
RN   [2]
RP   SEQUENCE OF 1-152 FROM N.A.
RX   MEDLINE=84223957; PubMed=6328658;
RA   Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA   Saule S., Martin P., Stehelin D.;
RT   "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT   new type of oncogene.";
RL   Science 224:1456-1459(1984).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC       ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC       IN CHICKENS.
CC   -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC       RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02006; AAA42394.1; ALT_INIT.
DR   EMBL; K01216; AAA42400.1; -.
DR   PIR; A00644; TVYUH.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   CONFLICT     29     29       R -> W (IN REF. 2).
FT   CONFLICT    140    140       S -> F (IN REF. 2).
FT   CONFLICT    146    146       I -> V (IN REF. 2).
SQ   SEQUENCE   604 AA;  67633 MW;  76EBCDD06745D609 CRC64;
```

```
  Query Match             23.4%;  Score 1592;  DB 1;  Length 604;
  Best Local Similarity   50.0%;  Pred. No. 2.4e-78;
  Matches  347;  Conservative   77;  Mismatches  136;  Indels  134;  Gaps   18;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
              |||: | | ||  ||:||  :   : : |: |    || |  |||  |     :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705
                |   || : || |:| :||:|:  |: ::||   :|| |:|||||| ||||||||||
Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKETELRKVKVLGSGAFGPFNNFTVSFWLRVPKVSASHLEV----LRENTSP 761
              |||| :|||||| :|||||||||||     |:   | :|:     : |    ||| |||
Db        118 APNQAHLRILKETEFKKVKVLGSGAFG----TIYKGLWIPEGEKVKIPVAIKELREATSP 173

Qy        762 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 821
              |||||||||||||| | :|:| ||||||||||||:|||||||||||::||::  :|||
Db        174 KANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQY 233

Qy        822 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 881
              |||||:|||||:|||: ||||||||||||||:| ||||||||||:||  || ||||:|
Db        234 LLNWCVQIAKGMNYLEERRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEG 293

Qy        882 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 941
              |||||||||||||| | :|||||||||||||||||||:|||||||| ||  :|||||||
Db        294 GKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERL 353

Qy        942 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPL 1000
              |||||||||||||||||||:: ||:||||::|||:||||| |::||| :| :   ||
Db        354 PQPPICTIDVYMIMVKCWMIDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPT 413

Qy       1001 DSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLT 1060
              || |||:|:|::|| |:|||:||||| ||||                | ||
Db        414 DSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-------- 449

Qy       1061 LGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSED 1115
                          |:||     | |  : ::  |         :  |  |  : | :|||| |
Db        450 ----------SRTPLLSSLSATSNNSATNCID--------RNGQGHPVREDSFVQRYSSD 491

Qy       1116 PTVPLPSET--DGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTL 1173
              ||     |:  ||::      | ||||||  : |: ||
```

```
Db        492 PTGNFLEESIDDGFL------PAPEYVNQ--LMPKKPSTAM------------------ 524

Qy       1174 SPGKNGVVKDVFAF------------------GGAVENPEYLTPQGGAAPQPHPPPAFS 1214
                  |   :: |                    ||:|||||         :  |
Db        525 ------VQNQIYNFISLTAISKLPMDSRYQNSHSTAVDNPEYL--------NTNQSPLAK 570

Qy       1215 PAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1248
                |::  || |    :             :||:|
Db        571 TVFESSPYWIQSGNHQ----------INLDNPDY 594


RESULT 15
ERBB_AVIEU
ID   ERBB_AVIEU     STANDARD;      PRT;   540 AA.
AC   P11273;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN   V-ERBB.
OS   Avian erythroblastosis virus (strain ts167).
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=103898;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87064458; PubMed=2878364;
RA   Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT   "A single amino acid substitution in v-erbB confers a thermolabile
RT   phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT   cells.";
RL   Mol. Cell. Biol. 6:1751-1759(1986).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13179; AAA42401.1; -.
DR   PIR; A25231; TVFVEB.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW   Glycoprotein; Phosphorylation.
FT   DOMAIN      132    399       PROTEIN KINASE.
FT   NP_BIND     138    146       ATP (BY SIMILARITY).
FT   BINDING     165    165       ATP (BY SIMILARITY).
FT   ACT_SITE    257    257       BY SIMILARITY.
FT   VARIANT     270    270       H -> D (IN THERMOLABILE V-ERBB).
SQ   SEQUENCE   540 AA;  60412 MW;  5B53297AA068B65D CRC64;

  Query Match             22.4%;  Score 1524;  DB 1;  Length 540;
  Best Local Similarity   52.8%;  Pred. No. 9.3e-75;
  Matches  329;  Conservative   70;  Mismatches  126;  Indels   98;  Gaps   16;

Qy        587 CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
              |||: | | ||  ||:||  :   : :||: |    || |  |||  |      :|||
Db          3 CAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58

Qy        647 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGA 705
                |   || : || |:| :||:|:  |: ::||   :|| |:|||||| ||||||||||

Db         59 NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRRLLQERELVEPLTPSGE 117

Qy        706 MPNQAQMRILKETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSP 761
              |||| :|||||| :|||||| ||||    ||  | :|:     : |    ||| |||
Db        118 APNQAHLRILKETEFKKVKVLGFGAFG----TVYKGLWIPEGEKVTIPVAIKELREATSP 173

Qy        762 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 821
              |||||||||||||| | :|:| ||||||||||||:|||||||||||::||::  :|||
Db        174 KANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQY 233

Qy        822 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 881
              ||||:|||||:|||:  :|||||||||||:| ||||||||||: |  || ||||:|
Db        234 LLNWCVQIAKGMNYLEERHMVHRDLAARNVLVKTPQHVKITDFGLAKQLGADEKEYHAEG 293

Qy        882 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 941
              |||||||||||||| | :|||||||||||||||||:|||||||| ||  :|||||||
Db        294 GKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERL 353

Qy        942 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPL 1000
              |||||||||||||||||| |:: ||:||||::|||:||||| |::||| :| :   ||
Db        354 PQPPICTIDVYMIMVKCWMSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMHLPSPT 413

Qy       1001 DSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLT 1060
              || |||:|:|::|| |:|||:||||| ||||                  | ||
Db        414 DSKFYRTLMEEEDMEDIVDADEYLVPHQGFF----------------NSPST-------- 449

Qy       1061 LGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSED 1115
                          |:||      | |  : ::   | : |              |
Db        450 ----------SRTPLLSSLSATSNNSATNCIDRNGG-----------H----------- 476

Qy       1116 PTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT-LERAKTLS 1174
                |    ||::      | ||||||  : |: ||                | : :    |
Db        477 -----PVREDGFL------PAPEYVNQ--LMPKKPSTAMVQNQIYNYISLTAISKLPIDS 523

Qy       1175 PGKNGVVKDVFAFGGAVENPEYL 1197
                :|       :   ||:|||||
Db        524 RYQN-------SHSTAVDNPEYL 539


Search completed: July 22, 2003, 09:19:00
Job time : 21.642 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.8246 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-710-730-14
Perfect score: 6813
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6091	89.4	1259	6	O18735	O18735 canis famil
2	3033	44.5	1209	11	Q9QX70	Q9qx70 rattus norv
3	3004	44.1	1210	11	Q9EP98	Q9ep98 mus musculu
4	2617	38.4	1165	13	Q9YH40	Q9yh40 xiphophorus
5	2580.5	37.9	1137	13	Q9W6F6	Q9w6f6 gallus gall
6	2204	32.3	1328	13	P79754	P79754 fugu rubrip
7	1923.5	28.2	1433	5	Q9BIH9	Q9bih9 anopheles g
8	1871	27.5	419	4	Q9UK79	Q9uk79 homo sapien
9	1739	25.5	367	11	Q8R2X1	Q8r2x1 mus musculu
10	1697.5	24.9	412	4	Q8WYV0	Q8wyv0 homo sapien
11	1609	23.6	729	15	Q86712	Q86712 avian rous-
12	1607	23.6	567	15	Q86714	Q86714 avian rous-
13	1547.5	22.7	962	15	Q64895	Q64895 avian eryth
14	1539	22.6	545	15	Q85468	Q85468 avian eryth
15	1506.5	22.1	655	11	Q9WVF5	Q9wvf5 mus musculu
16	1490.5	21.9	643	11	Q9ERV6	Q9erv6 mus musculu
17	1201	17.6	1193	5	Q9Y1X8	Q9y1x8 ephydatia f
18	1126	16.5	527	13	Q90836	Q90836 gallus gall
19	1117.5	16.4	1368	5	Q23821	Q23821 caenorhabdi
20	1114	16.4	1717	5	Q26566	Q26566 schistosoma
21	1001.5	14.7	478	11	Q9ESE0	Q9ese0 rattus norv
22	942.5	13.8	599	13	Q9PSH2	Q9psh2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	887	13.0	176	11	Q923V5	Q923v5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8szw1 drosophila
27	754.5	11.1	311	13	Q99162	Q99162 xiphophorus
28	734	10.8	331	4	Q9BUD7	Q9bud7 homo sapien
29	723	10.6	149	6	Q9BG66	Q9bg66 oryctolagus
30	695.5	10.2	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
31	679	10.0	1671	5	Q9NJV5	Q9njv5 biomphalari
32	655	9.6	1418	13	O93457	O93457 scophthalmu
33	646.5	9.5	1368	13	Q8UW85	Q8uw85 paralichthy
34	631.5	9.3	1369	13	Q8UW86	Q8uw86 paralichthy
35	630.5	9.3	1472	5	Q9U5A8	Q9u5a8 bombyx mori
36	621	9.1	1412	13	Q8UW84	Q8uw84 paralichthy
37	611	9.0	1358	13	O73798	O73798 xenopus lae
38	604.5	8.9	1418	13	Q8UW83	Q8uw83 paralichthy
39	592	8.7	1245	13	Q9YGH8	Q9ygh8 scophthalmu
40	583	8.6	1371	11	Q9QVW4	Q9qvw4 rattus sp.
41	580.5	8.5	2144	5	Q9VD94	Q9vd94 drosophila
42	578.5	8.5	1091	4	Q9UMQ4	Q9umq4 homo sapien
43	572	8.4	987	11	Q91YM0	Q91ym0 mus musculu
44	569	8.4	935	4	Q96L35	Q96l35 homo sapien
45	569	8.4	987	11	Q99MR2	Q99mr2 mus musculu

ALIGNMENTS

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErbB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB008451; BAA23127.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

```
  Query Match             89.4%;  Score 6091;  DB 6;  Length 1259;
  Best Local Similarity   89.2%;  Pred. No. 0;
  Matches 1128;  Conservative   49;  Mismatches   72;  Indels   16;  Gaps    5;

Qy          1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60
              ||||| |||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db          1 MELAAWCRWGLLLALLPSGAAGTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNL 60

Qy         61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
              |||||| ||||||||||||||||||||:||||:||||||||||||||||||||||||||
Db         61 ELTYLPANASLSFLQDIQEVQGYVLIAHSQVRQIPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy        121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
              |||     |  ||: |||||||||||||||||||||:||||:||||||||:|||||||
Db        121 DPLEGGIPAPGAAQGGLRELQLRSLTEILKGGVLIQRSPQLCHQDTILWKDVFHKNNQLA 180

Qy        181 LTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
              |||||||  ||  |||| ||  : |||  || ||||||||||||||||||| |||||||||
Db        181 LTLIDTNRFSACPPCSPACKDAHCWGASSGDCQSLTRTVCAGGCARCKGPQPTDCCHEQC 240

Qy        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300

Qy        301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGIT 360
              |||||||||||||||:|||||||||||||||||||||||||:::
Db        301 YNYLSTDVGSCTLVCPLNNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy        361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
              |||||||||||||||||||||||||||||||||||:||| ||||||||||||||||||
Db        361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLRVFEALEEITGYLYISAWPDSLP 420

Qy        421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTV 480
              :|||||:|||||:||:|||||||||||||||||||||||||| |  |||||||
Db        421 NLSVFQNLRVIRGRVLHDGAYSLTLQGLGISWLGLRSLRELGSGLALIHRNARLCFVHTV 480

Qy        481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
              |||||||||||||:|||||:|||||||||:  || ||||||||||||||||||||||||
Db        481 PWDQLFRNPHQALLHSANRPEEECVGEGLACYP-CAHGHCWGPGPTQCVNCSQFLRGQEC 539

Qy        541 VEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARC 600
              |||||||||||||||  |:|||||  |||||||||||| ||||||||||||||||||||
Db        540 VEECRVLQGLPREYVKDRYCLPCHSECQPQNGSVTCFGSEADQCVACAHYKDPPFCVARC 599

Qy        601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
              |||||||||:|||||| |||| |||||||||||| |||:|||||||||||:|||::||||
Db        600 PSGVKPDLSFMPIWKFADEEGTCQPCPINCTHSCADLDEKGCPAEQRASPVTSIIAAVVG 659

Qy        661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
              ||| ||:|:| |||||||:|||||||||||||||||||||||||||||||||||||||||
Db        660 ILLAVVVGLVLGILIKRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 719

Qy        721 RKVKVLGSGAFGFNNFTVSFWLRVP-----KVSASHLEVLRENTSPKANKEILDEAYVMA 775
              |||||||||||     ||   : :|     |:  : ::|||||||||||||||||||||
Db        720 RKVKVLGSGAFG----TVYKGIWIPDGENVKIPVA-IKVLRENTSPKANKEILDEAYVMA 774

Qy        776 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSY 835
              ||||||||||||||||||||||||||||||||||||:|||||||||||||:||||||||
Db        775 GVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVREHRGRLGSQDLLNWCVQIAKGMSY 834

Qy        836 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIL 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 LEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESIP 894

Qy        896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 PRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIM 954

Qy        956 VKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1015
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        955 VKCWMIDSECRPRFRELVAEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMG 1014

Qy       1016 DLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPL 1075
              |||||||||||||||||:| ||||| | ||||||||:|||:|||||||||||| |:|||
Db       1015 DLVDAEEYLVPQQGFFCPEPTPGAGGTAHRRHRSSSTRNGGGELTLGLEPSEEEPPKSPL 1074

Qy       1076 APSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
              |||||||||||||||||||||||||: |||||||||||||||| |||| |||||||||
Db       1075 APSEGAGSDVFDGDLGMGAAKGLQSLPSQDPSPLQRYSEDPTVPLPPETDGKVAPLTCSP 1134

Qy       1136 QPEYVNQPDVRPQPPSPREGPLPAARPAGATLER-----AKTLSPGKNGVVKDVFAFGGA 1190
              ||||||||:| ||||    ||||| :|||||||||      |||||||||||||||| |
Db       1135 QPEYVNQPEVWPQPPLALEGPLPPSRPAGATLERPKTLSPKTLSPGKNGVVKDVFAFGSA 1194

Qy       1191 VENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
              ||||||| |:| ||||||||||||||||||||||||| |||:|||||:||||||||||
Db       1195 VENPEYLAPRGRAAPQPHPPPAFSPAFDNLYYWDQDPSERGSPPSTFEGTPTAENPEYLG 1254

Qy       1251 LDVPV 1255
              |||||
Db       1255 LDVPV 1259


RESULT 2
Q9QX70
ID   Q9QX70      PRELIMINARY;      PRT;  1209 AA.
AC   Q9QX70;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor.
GN   EGFR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RX   MEDLINE=90258888; PubMed=2342466;
RA   Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA   Earp H.S.;
RT   "A truncated, secreted form of the epidermal growth factor receptor is
RT   encoded by an alternatively spliced transcript in normal rat tissue.";
RL   Mol. Cell. Biol. 10:2973-2982(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Petch L.A.;
RL   Submitted (NOV-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=LIVER;
RA   Guttridge K., Dawson T.L., Earp H.S.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M37394; AAF14008.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1209 AA;  134891 MW;  96FEE7F6CC1B7773 CRC64;

  Query Match             44.5%;  Score 3033;  DB 11;  Length 1209;
  Best Local Similarity   49.1%;  Pred. No. 3.7e-223;
```

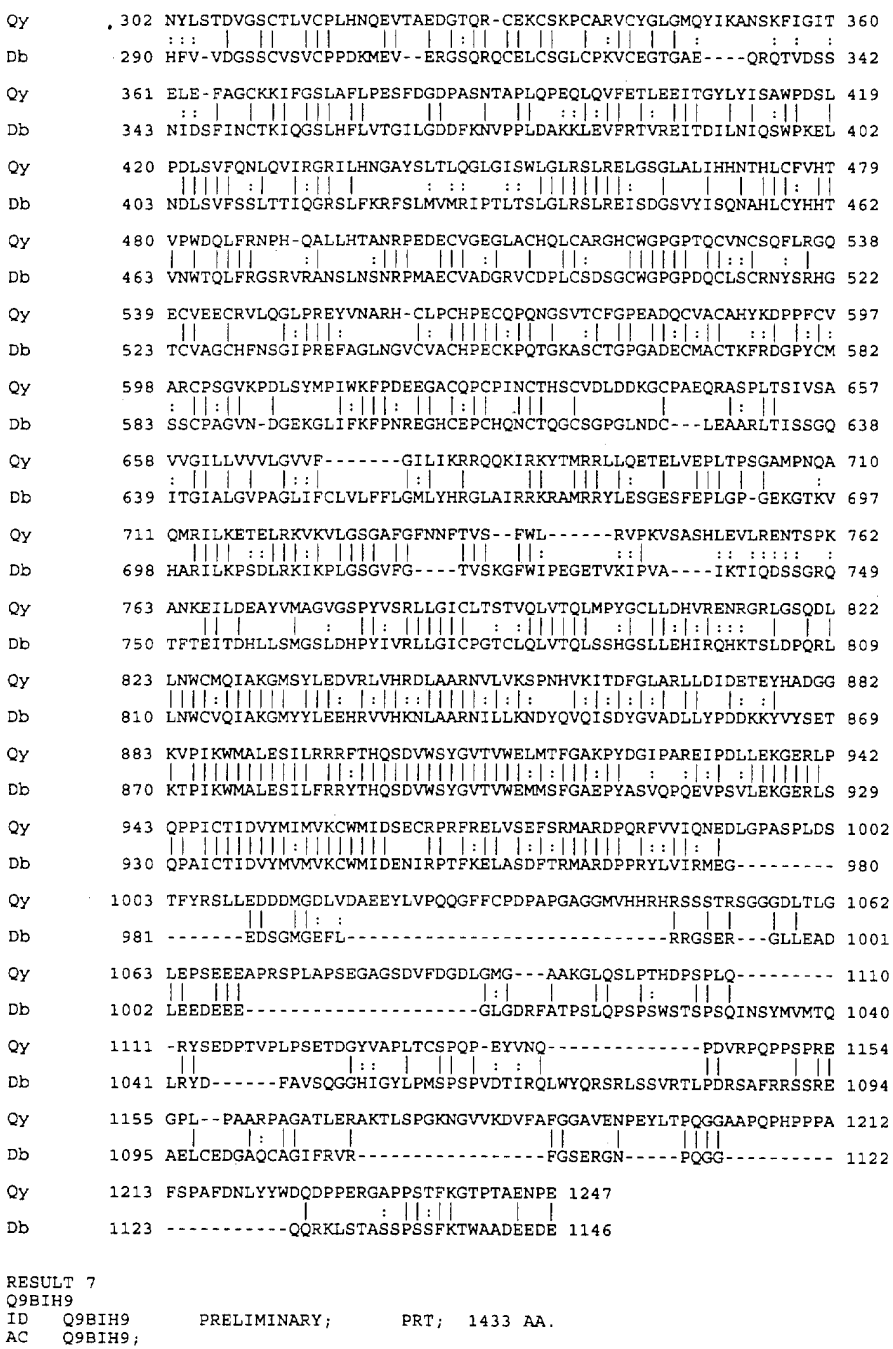
```
  Matches  630;  Conservative  166;  Mismatches  362;  Indels  124;  Gaps   30;

Qy          3 LAALCRWGLLLALLPPGA-ASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLE  61
              ||||| |         ||    :|| ||  :|    : | |   |: ::  |:|| ||||
Db         15 LAALCAAG--------GALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMFNNCEVVLGNLE  66

Qy         62 LTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
              :||:  |  ||||: |||| |||||| | | ::||: |:|:|| |:|: ||||||
Db         67 ITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAVLSN-- 124

Qy        122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLAL 181
                  | :  ||||| :|:| ||| | |    || ||  :|| |:||  ::  |:
Db        125 --------YGTNKTGLRELPMRNLQEILIGAVRFSNNPILCNMETIQWRDIV-QDVFLSN 175

Qy        182 TLIDTNRS-RACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQ 239
              :|  |    | | | |     |||    |:|| ||: :||  |: ||:|  |:|||| |
Db        176 MSMDVQRHLTGCPKCDPSCPNGSCWGRGEENCQKLTKIICAQQCSRRCRGRSPSDCCHNQ 235

Qy        240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
              |||||||: |||| |  |      |:  || |: ||  |::   ||||:|:|||:||  |
Db        236 CAAGCTGPRESDCLVCHRFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKC 295

Qy        300 PYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGI 359
              | ||: || |||   |      ||  ||| :|:||  || :|| |:|:  | ::  |
Db        296 PRNYVVTDHGSCVRACGPDYYEV-EEDGVSKCKKCDGPCRKVCNGIGIGEFK-DTLSINA 353

Qy        360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDS 418
              | :: |  |  | | |  || :| ||  : | || | :|:: :|::||||:| | |||::
Db        354 TNIKHPKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPEN 413

Qy        419 LPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVH 478
              ||  |:||::|||   :| :|| : || |: ||||||:|:  |  :|  | :||: :
Db        414 WTDLHAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYAN 473

Qy        479 TVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQ 538
              |: | :||  |:|      || | :|      |: ||:   |||| || ||:|    ||:
Db        474 TINWKKLFGTPNQKTKIMNNRAEKDCKATNHVCNPLCSSEGCWGPEPTDCVSCQNVSRGR 533

Qy        539 ECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVA 598
              |||::| :|:| |||:|    |: ||||| ||  ::|| |   | |: |||| | | ||
Db        534 ECVDKCNILEGEPREFVENSECIQCHPECLPQTMNITCTGRGPDNCIKCAHYVDGPHCVK 593

Qy        599 RCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASP-LTSIVSA 657
              ||||:  : : : :||| |    |  |  |||: |     |||  :|   | : || :
Db        594 TCPSGIMGENNTL-VWKFADANNVCHLCHANCTYGCAGPGLKGC--QQPEGPKIPSIATG 650

Qy        658 VVGILLVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
              :|| || :|: |  || |  ||:| :|| |:|||||| |||||||||  |||| :||||
Db        651 IVGGLLFIVV-VALGIGLFMRRRQLVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILK 709

Qy        717 ETELRKVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSPKANKEILDEAY 772
              ||| :|:|||||||||    ||   | :|:     : |    ||| |||||||||||||
Db        710 ETEFKKIKVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAY 765

Qy        773 VMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKG 832
              ||| | :|:| |||||||||||||:|||||||||||:|||::  :||| |||||:|||||
Db        766 VMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKG 825

Qy        833 MSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALE 892
              |:|||| |||||||||||||:| ||||||||||:||  :| ||||:|||||||||||
Db        826 MNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALE 885

Qy        893 SILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVY 952
              ||| | :|||||||||||||||||||:|||||||| ||  :|||||||||||||||||||
Db        886 SILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVY 945

Qy        953 MIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLED 1011
              |||||||||:: ||:|||| ||:|||||||::||| :| :   || || |||:|:|:
Db        946 MIMVKCWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMEE 1005

Qy       1012 DDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAP 1071
              :|| |:|||:|||:|||||                | ||
Db       1006 EDMEDVVDADEYLIPQQGFF---------------NSPST-----------------S 1031

Qy       1072 RSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVA 1129
              |:||  |  | |:       :        |    : : ||||| |||  |  :   | ::
Db       1032 RTPLLSSLSANSN----SSTVACINRNGSCRVKEDAFLQRYSSDPTSVLTEDNIDDTFL- 1086

Qy       1130 PLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGG 1189
                    | |||:||  |  :|    : |:   :|           :||::   ::
Db       1087 -----PVPEYINQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSN 1130

Qy       1190 AVENPEYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGA 1232
              || ||||| | |        |   |  ||:   | |          :|       |:
Db       1131 AVSNPEYLNTAQ---------PTCLSSGFDSSALWIQKGSHQMSLDNPDYQQDFFPKEAK 1181

Qy       1233 PPSTFKGTPTAENPEYLGLDVP 1254
              |   ||| ||||| ||| :  |
Db       1182 PNGIFKG-PTAENAEYLRVAPP 1202
```

RESULT 3
Q9EP98
ID   Q9EP98   PRELIMINARY;   PRT; 1210 AA.
AC   Q9EP98;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Epidermal growth factor receptor isoform 1.
GN   EGFR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative Egfr transcripts encoding truncated receptor
RT   isoforms.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA   Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA   Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA   Maihle N.J.;
RT   "Comparative genomic sequence analysis and isolation of human and
RT   mouse alternative Egfr transcripts encoding truncated receptor
RT   isoforms.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF275366; AAG28045.1; -.
DR   EMBL; AF275364; AAG28045.1; JOINED.
DR   EMBL; AF275365; AAG28045.1; JOINED.
DR   EMBL; AF275367; AAG24386.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95294; Egfr.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 5.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.

```
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Receptor; Transferase.
SQ   SEQUENCE   1210 AA;  134840 MW;  62CD021C9DE32E18 CRC64;

  Query Match           44.1%;  Score 3004;  DB 11;  Length 1210;
  Best Local Similarity  48.6%;  Pred. No. 6.2e-221;
  Matches  621;  Conservative  168;  Mismatches  368;  Indels  120;  Gaps   27;

Qy         11 LLLALLPPGAA--STQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTN 68
Db         14 LLTALCAAGGALEEKKVCQGTSNRLTQLGTFEDHFLSLQRMYNNCEVVLGNLEITYVQRN 73

Qy         69 ASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db         74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNALYENTYALAILSN--------- 124

Qy        129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHKNNQLALTLI 184
Db        125 -YGTNRTGLRELPMRNLQEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180

Qy        185 DTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAG 243
Db        181 -QSHPSSCPKCDPSCPNGSCWGGGEENCQKLTKIICAQQCSHRCRGRSPSDCCHNQCAAG 239

Qy        244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db        240 CTGPRESDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNY 299

Qy        304 LSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMQYIKANSKFIGITELE 363
Db        300 VVTDHGSCVRACGPDYYEV-EEDGIRKCKKCDGPCRKVCNGIGIGEFK-DTLSINATNIK 357

Qy        364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDL 422
Db        358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

Qy        423 SVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPW 482
Db        418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGDVIISGNRNLCYANTINW 477

Qy        483 DQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE 542
Db        478 KKLFGTPNQKTKIMNNRAEKDCKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGRECVE 537

Qy        543 ECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAHYKDPPFCVARCPS 602
Db        538 KCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy        603 GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVGIL 662
Db        598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCEVWPSGPKIPSIATGIVGGL 656

Qy        663 LVVVLGVVFGI-LIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELR 721
Db        657 LFIVV-VALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQAHLRILKETEFK 715

Qy        722 KVKVLGSGAFGFNNFTVSFWLRVPKVSASHLEV----LRENTSPKANKEILDEAYVMAGV 777
Db        716 KIKVLGSGAFG----TVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASV 771

Qy        778 GSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLE 837
Db        772 DNPHVCRLLGICLTSTVQLITQLMPYGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLE 831

Qy        838 DVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRR 897
Db        832 DRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHR 891

Qy        898 RFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVK 957
Db        892 IYTHQSDVWSYGVTVWELMTFGSKPYDGIPASDISSILEKGERLPQPPICTIDVYMIMVK 951

Qy        958 CWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMGD 1016
Db        952 CWMIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMED 1011

Qy       1017 LVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA 1076
Db       1012 VVDADEYLTPQQGFF----------------NSPST----------------SRTPLL 1037

Qy       1077 PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCS 1134
Db       1038 SSLSATSN----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDDAFL------ 1087

Qy       1135 PQPEYVNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENP 1194
Db       1088 PVPEYVNQ-SVPKRPAGSVQNPVYHNQPLHP--------APGRDLHYQN--PHSNAVGNP 1136

Qy       1195 EYL-TPQGGAAPQPHPPPAFSPAFDNLYYWDQ---------DP-------PERGAPPSTF 1237
Db       1137 EYLNTAQ---------PTCLSSGFNSPALWIQKGSHQMSLDNPDYQQDFFPKETKPNGIF 1187

Qy       1238 KGTPTAENPEYLGLDVP 1254
Db       1188 KG-PTAENAEYLRVAPP 1203


RESULT 4
Q9YH40
ID   Q9YH40      PRELIMINARY;      PRT;  1165 AA.
AC   Q9YH40;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase proto-oncogene.
GN   XMRK.
OS   Xiphophorus xiphidium.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX   NCBI_TaxID=8086;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RX   MEDLINE=98241172; PubMed=9582016;
RA   Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA   Altschmied J., Schartl M.;
RT   "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT   overexpression and mutational alterations.";
RL   Oncogene 16:1681-1690(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO PURIFICATION;
RA   Schartl M.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53471; AAD10500.2; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
```

```
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1165 AA;  129614 MW;  7F7EE38D8771A74E CRC64;

  Query Match             38.4%;  Score 2617;  DB 13;  Length 1165;
  Best Local Similarity   44.5%;  Pred. No. 2.6e-191;
  Matches  571;  Conservative  164;  Mismatches  385;  Indels  164;  Gaps   35;

Qy          1 MELAALCRWGLLLALLPPG-AAST----QVCTGTDMKLRLPASPETHLDMLRHLYQGCQV 55
              :||  |    ||| ||  |   ||    :|| ||  :: :   : |   :: :| || |
Db          4 LELLEL----LLLLLLSIGRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKKMYSGCNV 56

Qy         56 VQGNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALA 115
              |  |||:||    |  ||||| |||| |||||| |:|  :|| ||::||  |:| |: |
Db         57 VLENLEITYTQENQDLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLYEGNFTLL 116

Qy        116 VLDNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHK 175
              |: |    | ::|       ||::||| :||||| ||| :  || ||  :|| | ||  |
Db        117 VMSNYQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHNPLLCNVETINWWDIVDK 173

Qy        176 NNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTD 234
              :  : ||     | | | | |     ||     ||   |: :||  |  ||:|| | |
Db        174 TSNPTMNLIPHAFERQCQKCDPGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPID 233

Qy        235 CCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
              ||:| || ||||| :|||||  ||  | |:  ||     |:  : : : ||  :|||||:
Db        234 CCNEHCAGGCTGPRATDCLACRDFNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAA 293

Qy        295 CVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGM----QYI 350
              ||  || ||: |: |:|    |      ||  |:| : |: |   | :|| |:|:      |
Db        294 CVKECPSNYVVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGIGSLSNTI 351

Qy        351 KANSKFIGITELEFAGCKKIPGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
                 ||  ||      |: | || | :      ||:|||      |: || |    |::|||||
Db        352 AVNSTNIG----SFSNCTKINGDIILNRNSFEGDPHYKIGPMDPEHLWNLTTVKEITGYL 407

Qy        411 YISAWPDSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGLRSLRELGSGLALIH 469
              |  ||:::  |||||||::||||    :  :| : :|   : |||| ||:|: :|  ::
Db        408 VIMWWPENMTSLSVFQNLEIIRGRTTFSRGFSFVVVQVSHLQWLGLSSLKEVSAGNVILK 467

Qy        470 HNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCV 529
              :   | :  |: | :|||:  |:: : |          |   |:  |:   |||||| ||
Db        468 NTPQLRYASTINWRRLFRSEDQSIEYDART-------ENQTCNNECSEDGCWGPGPTMCV 520

Qy        530 NCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVACAH 589
              :|    ||  ||  | :||| |||       |: || ||  |  |:||:||    | |||
Db        521 SCLHVDRGGRCVASCNLLQGEPREAQVDGRCVQCHQECLVQTDSLTCYGPGPANCSKCAH 580

Qy        590 YKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS 649
              ::|  | |: ||| |:  |   : |||: |: | |||  |||  |       ||  :   |
Db        581 FQDGPQCIPRCPHGMLGDGDTL-IWKYADKMGQCQPCHQNCTQGCSGPGLSGCRGD-IVS 638

Qy        650 PLTSIVSAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQ 709
              :  |  | |:|: |:: ::  :|::||: | || |:|||||  |||||||||  |||
Db        639 HSSLAVGLVSGLLITVIVALLIVVLLRRRRIK-RKRTIRRLLQEKELVEPLTPSGQAPNQ 697

Qy        710 AQMRILKETELRKVKVLGSGAFGFNNFTV--SFW------LRVPKVSASHLEVLRENTSP 761
              | :|||||| :| :|||||||    ||    |      :|:|      ::|||| |||
Db        698 AFLRILKETEFKKDRVLGSGAFG----TVYKGLWNPDGENIRIPVA----IKVLREATSP 749

Qy        762 KANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQD 821
              | |:|:|||||||  |  |:| |||||||||  |||||||||||||||:||::: |:  |
Db        750 KVNQEVLDEAYVMASVDHPHVCRLLGICLTSAVQLVTQLMPYGCLLDYVRQHQERICGQW 809

Qy        822 LLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADG 881
              |||||:|||||:|||:  |||||||||||:|:||||||||||::||  || || | |
Db        810 LLNWCVQIAKGMNYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQAHG 869

Qy        882 GKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERL 941
              |||||||||||||:  :|||||||||||||||||||:||||||||:||  :|| ||||
Db        870 GKVPIKWMALESILQWTYTHQSDVWSYGVTVWELMTFGSKPYDGIPAKEIASVLENGERL 929

Qy        942 PQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQNEDLGPASPLD 1001
              ||||||||:||||::||||||   |||||||| |||:|||||  |::|||       || |
Db        930 PQPPICTIEVYMIILKCWMIDPSSRPRFRELVGEFSQMARDPSRYLVIQG---NLPSPSD 986

Qy       1002 STFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTL 1061
                :  ||  ||  |:|||:|||:                        |::  : | |
Db        987 RRLFSRLLSSDD--DVVDADEYLL-------------------RYKRIN-RQGS----- 1018

Qy       1062 GLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTV-PL 1120
                        |  |  |                       |  : |  ||  |||   |
Db       1019 -----------EPCIPPNGH-------------------PVRENSIALRYISDPTQNAL 1047

Qy       1121 PSETDGYVAPLTCSPQPEYVNQPDVRPQP-------PSPRE-----GPLP-AARPAGATL 1167
                : ||:          ||||||               |:  :    ||:  ::: |
Db       1048 EKDLDGH----------EYVNQPGSETSSRLSDIYNPNYEDLTDGWGPVSLSSQEAETNF 1097

Qy       1168 ERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDP 1227
                | : |:  :| :         |::::|:|   | |           :  ||
Db       1098 SRPEYLNTNQNSL---PLVSSGSMDDPDY---QAG----------YQAAF---------L 1132

Qy       1228 PERGAPPSTFKGTPTAENPEYLGL 1251
              |: ||        | ||| |||||
Db       1133 PQTGALTGNGMFLPAAENLEYLGL 1156


RESULT 5
Q9W6F6
ID   Q9W6F6      PRELIMINARY;      PRT;  1137 AA.
AC   Q9W6F6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Receptor tyrosine kinase (Fragment).
GN   ERBB4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HINDBRAIN;
RX   MEDLINE=99263203; PubMed=10328884;
RA   Dixon M., Lumsden A.;
RT   "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
RT   embryonic chick hindbrain.";
RL   Mol. Cell. Neurosci. 13:237-258(1999).
DR   EMBL; AF121963; AAD31764.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001368; TNFR_c6.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 1.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW   Kinase; Tyrosine-protein kinase.
FT   NON_TER       1      1
SQ   SEQUENCE   1137 AA;  127927 MW;  4D616436F87DC84F CRC64;

  Query Match             37.9%;  Score 2580.5;  DB 13;  Length 1137;
```

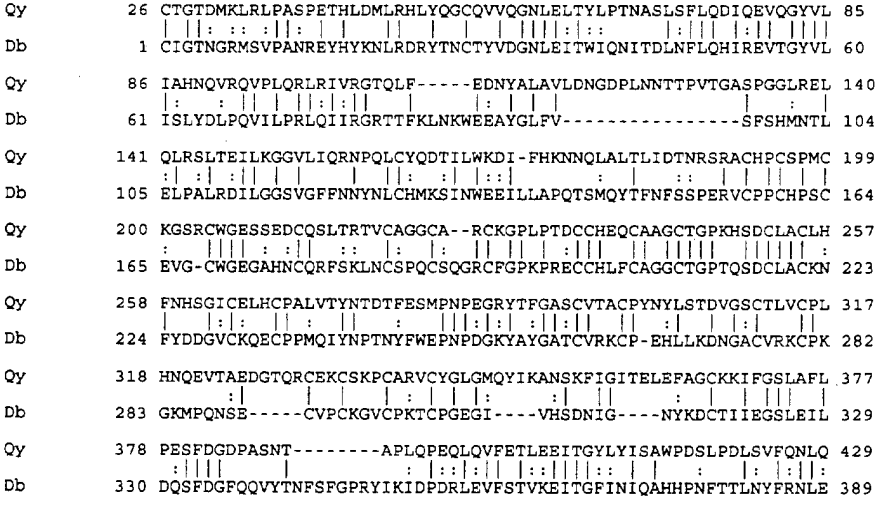
```
  Best Local Similarity   45.5%;  Pred. No. 1.6e-188;
  Matches  521;  Conservative  177;  Mismatches  358;  Indels   89;  Gaps   29;

Qy        161 LCYQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVC 220
              ||: ||| |:|| ||: || | | | | | | |||| :  ||:||:|||
Db          3 LCFADTIHWQDIVRNPWASNFTLVPTNGSSGCGRCHKSCTG-RCWGPTENHCQTLTKTVC 61

Qy        221 AGGC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTF 279
              |  |  || ||  :|||| :|| ||:||| :|| ||::|| || |  ||   ||  ||
Db         62 AEQCDGRCYGPYVSDCCHRECAGGCSGPKDTDCFACMNFNDSGACVTQCPQTFVYNPTTF 121

Qy        280 ESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCA 339
              :   |   :||:|| ||  ||:|::  |  ||   ||    ||  |:| : |: |:  |
Db        122 QLEHNHNAKYTYGAFCVKKCPHNFV-VDSSSCVRACPSSKMEV-EENGIKMCKPCTDICP 179

Qy        340 RVCYGLGMQYIKANSKFIGITELE-FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQ 398
              : | |:|   :  ::: :  : :: |  | || |:| ||     |||     | : ||:|
Db        180 KACDGIGTGSL-VSAQTVDSSNIDKFINCTKINGNLIFLVTGIHGDPYHTIAAINPEKLN 238

Qy        399 VFETLEEITGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSL 458
              :|:|: |||||| | :||::: |  || ||  | || |::|   | |:  ||: |  :||
Db        239 IFQTVREITGYLNIQSWPENMTDFRVFSNLVTIGGRALYSGLSLLILKQQGITSLQFQSL 298

Qy        459 RELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARG 518
              ::: :|   |  |::||: ||| |  ||  | |  :   |:  : |  :|: |::||:
Db        299 KQISAGNIYITDNSNLCYYHTVNWTSLFSTPSQKTVIHRNKKAENCTADGMVCNELCSSD 358

Qy        519 HCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQP-QNGSVTCF 577
              |||||| ||::| :|:||: |:| | :  |  ||: |   |: | |:|:  ::  :||:
Db        359 GCWGPGPDQCLSCKRFIRGRTCIESCNLYDGEFREFANGSVCMECDPQCEKMEDNMITCY 418

Qy        578 GPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
              ||  | |  | |:|| | || :|| |::   |:  |:|: ||:  | ||  |||  |
Db        419 GPGPDHCTKCFHFKDGPNCVEKCPDGLQGANSF--IFKYADEDRECHPCHPNCTQGCRGP 476

Qy        638 DDKGC-----------PAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQQKIRKY 685
                  |           |.  | :||  | : |: |: ::|::|: | : ::|:  | :|
Db        477 ASHDCIYYPWTRQSTLPQHAR-TPL--IAAGVIGGLFIIVIMGLTFAVYVRRKSIK-KKR 532

Qy        686 TMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGFNNFTVSFWLRVP 745
              :|| | |||||||||||||  |||||:||||||||::||||||||||    ||   : ||
Db        533 ALRRFL-ETELVEPLTPSGTAPNQAQLRILKETELKRVKVLGSGAFG----TVYKGIWVP 587

Qy        746 -----KVSASHLEVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQL 800
                  |:  : :::| | | |||| | :||| :|| :  |:: ||||:||: |:|||||
Db        588 EGETVKIPVA-IKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQL 646

Qy        801 MPYGCLLDHVRENRGRLGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVK 860
              ||:|||||:| |::  :||| |||||:|||||| ||| :|||||||||||||||||||||
Db        647 MPHGCLLDYVHEHKDNIGSQLLLNWCVQIAKGMMYLEERRLVHRDLAARNVLVKSPNHVK 706

Qy        861 ITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGA 920
              |||||||||: || ||:|||||:|||||||| |  |:|||||||||||||:|||||||
Db        707 ITDFGLARLLEGDEKEYNADGGKMPIKWMALECIHYRKFTHQSDVWSYGVTIWELMTFGG 766

Qy        921 KPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMA 980
              ||||||| |||||||||||||||||||||||||:|||||||:: ||:|:|| :||||||
Db        767 KPYDGIPTREIPDLLEKGERLPQPPICTIDVYMVMVKCWMIDADSRPKFKELAAEFSRMA 826

Qy        981 RDPQRFVVIQNED-LGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGA 1039
              |||||::||| :| :   || || |:::||:::|: |::|||||||| | |  | |   :
Db        827 RDPQRYLVIQGDDRMKLPSPNDSKFFQNLLDEEDLEDMMDAEEYLVP-QAFNIPPPIYTS 885

Qy       1040 GGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRS--PLAP-SEGAGSDVFDGDLGMGAAK 1096
                :     | ||     |: |    ||    | || ::|| :::|:     |  :
Db        886 RTRIDSNRNQFVYRDGGYAAEQGV-PMPYRAPGCIIPEAPVAQGATAEIFEDTCCNGTLR 944

Qy       1097 GLQSLPTHDPSPLQRYSEDPTVPLPS-------ETDGYVAPLTCSPQPEYVNQPDVRPQP 1149
                :    : |  |||| |||| :|        : |||: |:   |: :|:|  :  |
Db        945 KQVATLAKEDSSTQRYSADPTVFIPERVIRGELDEDGYMTPMRDKPKTDYLNPVEENPFV 1004

Qy       1150 PSPREGPLPAA-RPAGATLERAKTLSPGKNGVVKDVF--------AFGGAVENPEYLTPQ 1200
                : | | |   |            :|      :| :         |   :|| |||

Db       1005 SRRKNGDLQAVDNPEYHN-------APNGQPKAEDEYVNEPLYLNTFANTLENAEYL--- 1054

Qy       1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPERGA--PPSTFKGTPT--------------AE 1244
                    : : |     ||||  ||:   | |   |  :  |                 ||
Db       1055 -----KNNLPEKAKKAFDNPDYWNHSLPPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAE 1109

Qy       1245 NPEYL 1249
                |||||
Db       1110 NPEYL 1114


RESULT 6
P79754
ID   P79754      PRELIMINARY;      PRT;  1328 AA.
AC   P79754;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ErbB3.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99177347; PubMed=10077531;
RA   Gellner K., Brenner S.;
RT   "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT   rubripes.";
RL   Genome Res. 9:251-258(1999).
DR   EMBL; AF056116; AAC34391.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   1328 AA;  148613 MW;  A333039258B647E9 CRC64;

  Query Match             32.3%;  Score 2204;  DB 13;  Length 1328;
  Best Local Similarity   39.2%;  Pred. No. 1.4e-159;
  Matches  508;  Conservative  161;  Mismatches  418;  Indels  208;  Gaps   36;

Qy          9 WGLLLALLPP--GAASTQ----VCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLEL 62
              | |:|  :     |||:|    || ||   | | |   :: :  |:||::: ||||:
Db          4 WRLILMCVASRLRAASSQTQEAVCPGTQNGLSSTGSQENQYNLNKDRYKGCEIIMGNLEI 63

Qy         63 TYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDP 122
              | : :|   |||: |:|| |||||| |  :::|| :||::|| |:|  :||:|  |
Db         64 TQIESNWDFSFLKTIREVTGYVLIAMNHFQEIPLGQLRVIRGNSLYERRFALSVFLN--- 120

Qy        123 LNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALT 182
                  |  | | || :| | :||||| ||| |  |  | |   : |:||  :||   :
Db        121 ----YPKDG--PSGLNQLGLMNLTEILDGGVQIINNKYLRYGPWVYWRDII-RNNDAPIE 173

Qy        183 LIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGC-ARCKGPLPTDCCHEQCA 241
              :         ||      | |: |||   : || ||:||||  |  || |  | |||| :||
Db        174 IQFNGERGVCH---KSC-GNYCWGPGKDQCQILTKTVCAPQCNDRCFGTSPRDCCHIECA 229

Qy        242 AGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY 301
              ||| ||  :|| ||  || || |   ||  : ||  ||:   ||  :| :|: ||: ||
Db        230 AGCKGPLDTDCFACRLFNDSGACVPQCPQTLIYNKQTFQMETNPNAKYQYGSICVSQCPT 289
```

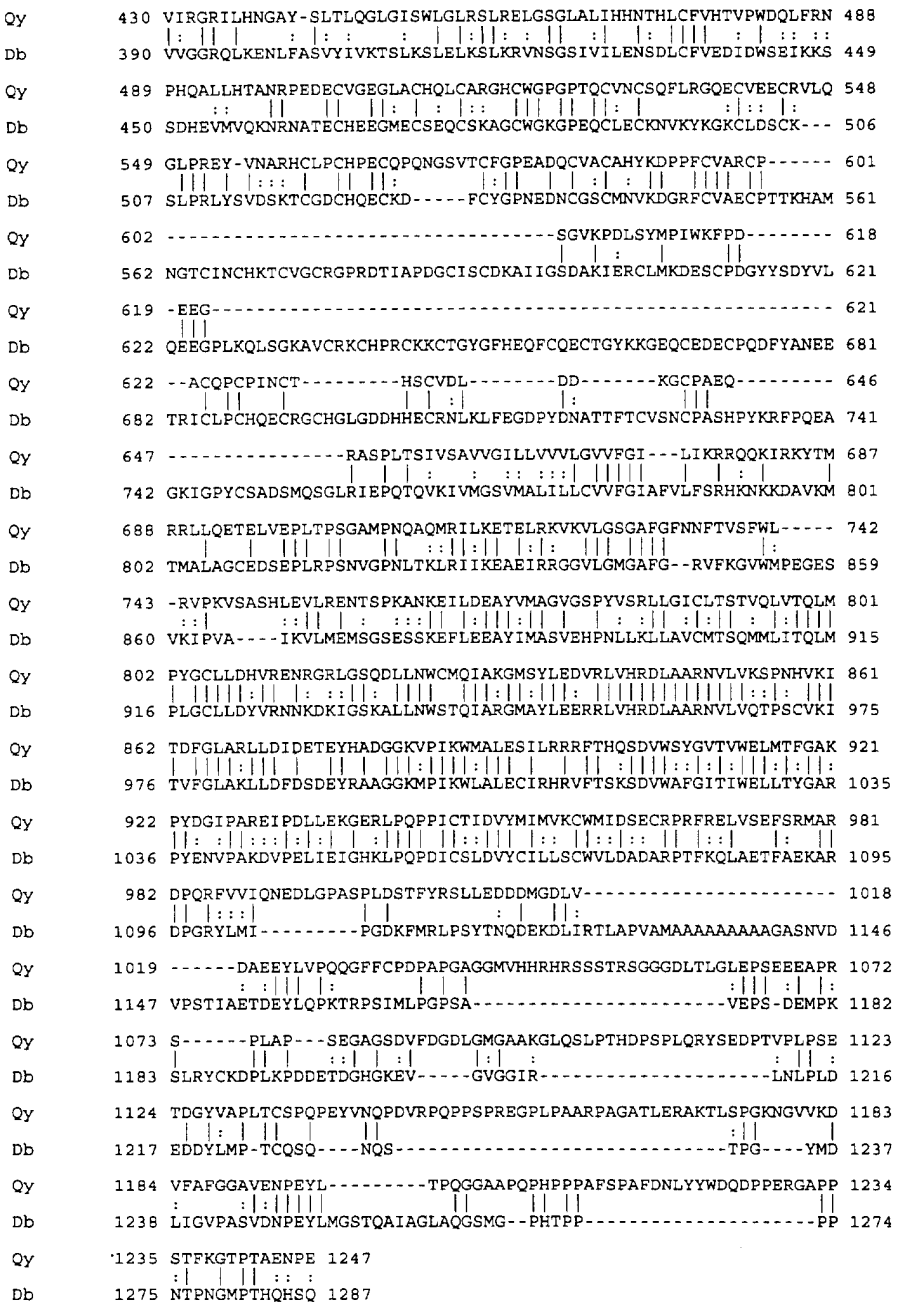


RESULT 7
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

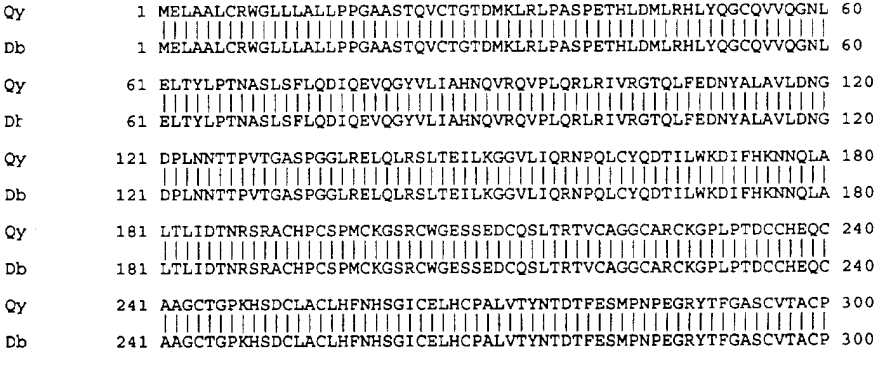
Query Match 28.2%; Score 1923.5; DB 5; Length 1433;
Best Local Similarity 31.6%; Pred. No. 4.6e-138;
Matches 459; Conservative 197; Mismatches 400; Indels 397; Gaps 40;





RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 7.7e-135;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;



RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

A DOCPHOENIX

BACKFILE DOCUMENT INDEX SHEET

APPL PARTS

_____ IMIS _____
Internal Misc. Paper

_____ LET. _____
Misc. Incoming Letter

_____ 371P _____
PCT Papers in a 371Application

_____ A... _____
Amendment Including Elections

_____ ABST _____
Abstract

_____ ADS _____
Application Data Sheet

_____ AF/D _____
Affidavit or Exhibit Received

_____ APPENDIX _____
Appendix

_____ ARTIFACT _____
Artifact

_____ BIB _____
Bib Data Sheet

_____ CLM _____
Claim

_____ COMPUTER _____
Computer Program Listing

_____ CRFL _____
All CRF Papers for Backfile

_____ DIST _____
Terminal Disclaimer Filed

_____ DRW _____
Drawings

_____ FOR _____
Foreign Reference

_____ FRPR _____
Foreign Priority Papers

_____ IDS _____
IDS Including 1449

_____ NPL _____
Non-Patent Literature

_____ OATH _____
Oath or Declaration

_____ PET. _____
Petition

_____ RETMAIL _____
Mail Returned by USPS

_____ SEQLIST _____
Sequence Listing

_____ SPEC _____
Specification

_____ SPEC NO _____
Specification Not in English

_____ TRNA _____
Transmittal New Application

_____ CTNF _____
Count Non-Final

_____ CTRS _____
Count Restriction

_____ EXIN _____
Examiner Interview

_____ M903 _____
DO/EO Acceptance

_____ M905 _____
DO/EO Missing Requirement

_____ NFDR _____
Formal Drawing Required

_____ NOA _____
Notice of Allowance

_____ PETDEC _____
Petition Decision

OUTGOING

_____ CTMS _____
Misc. Office Action

_____ 1449 _____
Signed 1449

_____ 892 _____
892

_____ ABN _____
Abandonment

_____ APDEC _____
Board of Appeals Decision

_____ APEA _____
Examiner Answer

_____ CTAV _____
Count Advisory Action

_____ CTEQ _____
Count Ex parte Quayle

_____ CTFR _____
Count Final Rejection

INCOMING

_____ AP.B _____
Appeal Brief

_____ C.AD _____
Change of Address

_____ N/AP _____
Notice of Appeal

_____ PA.. _____
Change in Power of Attorney

_____ REM _____
Applicant Remarks in Amendment

_____ XT/ _____
Extension of Time filed separate

Internal

_____ SRNT _____
Examiner Search Notes

_____ CLMPTO _____
PTO Prepared Complete Claim Set

_____ ECBOX _____
Evidence Copy Box Identification

_____ WCLM _____
Claim Worksheet

_____ WFEE _____
Fee Worksheet

File Wrapper

_____ FWCLM _____
File Wrapper Claim

_____ IIFW _____
File Wrapper Issue Information

_____ SRFW _____
File Wrapper Search Info

REV 7/02/03